THE UNIVERSITY OF WARWICK

MOLECULAR ORGANISATION AND ASSEMBLY IN CELLS DOCTORAL TRAINING CENTRE

Biochemical and Biophysical Characterisation of the Quartromicin Polyketide Synthase

---

Thesis

Submitted in partial requirement for the degree of

Doctor of Philiosophy

Author:

Richard John GIBSON

Supervisors:

Prof. Gregory CHALLIS

Dr. Józef LEWANDOWSKI

Dr. David ROPER

September 27, 2018

Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank my supervisors; Prof. Greg Challis for providing in-depth knowledge of PKS; Dr. Józef Lewandowski for providing invaluable NMR experience; Dr. David Roper for providing crystallography expertise. A special thanks should be given to both Greg and Józef, who provided unwavering levels of enthusiasm and patience that proved vital in my transition from mathematics to chemical biology, as well as taking the time to meet with me on a regular basis and providing valuable input and suggestions regarding my work.

My examiners, Prof. Dominic Campopiano and Dr. Claudia Blindauer, provided great feedback and allowed me to improve my thesis. I'd like to thank them both for the time they spent reading my thesis and for asking interesting questions regarding my work.

I would also like to thank both current and past members and staff of the chemical biology research facilities, especially the Challis group, who taught me important techniques and provided the much of previous work from which my project was developed (Dr. Orestis Lazos and Dr. Malek Zerikly). In particular, special thanks is due to Dr. Matthew Jenner, Dr. Lona Alkhalaf, Dr. Simone Kosol and Dr. Daniel Griffiths. Daniel supervised the MSc mini-projects that kindled my interest in PKS and Simone took the time to meet with me and discuss my project. Both also provided help in acquiring HSQCs. Matthew provided plenty of guidance and technical assistance in running mass spec samples, as well as providing the inspiration for much of the work in this thesis. Lona provided great assistance with setting up crystal trials and helping to structure my thesis. The considerable time she spent proofreading my entire thesis is much appreciated. Both Matthew and Lona provided plenty of great ideas and helped greatly at progress meetings. Dr. Chris Fage also provided technical help in setting up crystal trials.

Furthermore, I'd like to thank Matthew Beech, Christian Hobson, Dr. Shan Shan Zhou and Dr. Joleen Masschelein who provided substrates for extension studies. A special thank you goes to Matthew Beech for synthesising the decanoyl and dodecanoyl pantetheine substrates upon request, allowing me to probe the effect of chain length on substrate tolerance of M5. I'd also like to thank Dr. Richard Napier and Veselina Uzunova for providing great SPR-related advice and helping me learn how to conduct and analyse experiments using Biacore instruments.

Finally, I'd like to thank my fiancée, Micah Alyse Maria Conchita Josefina Pua Pastrana, for her love and support throughout my PhD.

Richard Gibson is an ESPRC funded student at the Molecular Organisation and Assembly in Cells DTC, the University of Warwick (Grant number EP/F500378/1).
Email: R.J.Gibson@Warwick.ac.uk

Declaration

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree.

The work presented (including data generated and data analysis) was carried out by the author except in the cases outlined below:

- Negative stain images (Ian Hands-Portman, from provided samples)
- MaXis LCMS-TQOF samples were run by Dr. Matthew Jenner, Dr. Joleen Masschelein, Dr. Lijiang Song and Marianne Costa
- Synthesis of SNAC and pantetheine-bound substrates (Matthew Beech, Christian Hobson, Dr. Zhou Shan Shan and Dr. Joleen Masschelein)
- Cryo-EM images and 2-D reconstructions (Dr. Daniel Griffiths)
- Carbene footprinting (Dr. Matthew Jenner).

Abstract

Quartromicin (QMN), derived from *Amycolatopsis orientalis* Q427-8, are a complex of novel antiviral antibiotics active against HSV-1, HIV and influenza. Their biosynthesis involves a type I polyketide synthase (PKS), proposed to implement a module skipping strategy that produces two different precursors (Figure 1). A series of Diels-Alderase reactions, followed by oxidation and glycosylation, leads to QMN biosynthesis. Determining how this module skipping mechanism works would greatly increase the understanding of how modules within PKSs can be relocated to facilitate PKS engineering.

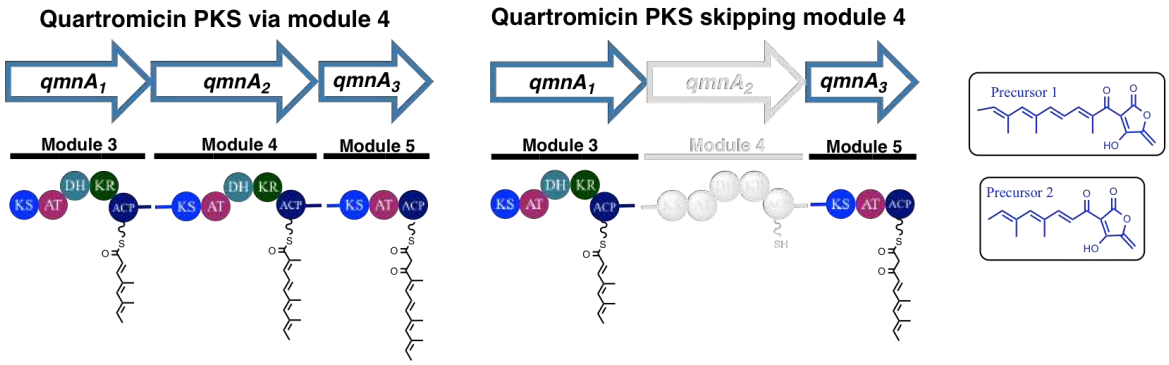


Figure 1. *The final three modules of the biosynthetic PKS pathway for quartromicin PKS and the resulting precursors.*

It was shown that module 5 (M5) can accept substrates from the upstream modules' acyl carrier protein (ACP) domains (M3_ACP and M4_ACP) and utilise malonyl-CoA to catalyse substrate extension. Substrate specificity did not appear to be a factor in maintaining pathway fidelity. M3_ACP was observed to have a quicker rate of acyl transfer to M5 than M4_ACP and stronger protein-protein interactions (K_D: 1.3 v 4.0 μM). Removal of M3_ACP's C-terminal docking domain resulted in slower transfer and weakening of protein-protein interactions (K_D: 2.6 μM). There was no significant change in rates of transfer or strength of interactions when the docking domain was omitted from M4_ACP. Sequence analysis of the docking domains highlighted the omission of a widely conserved negative residue from M4's ACP docking domain, perhaps explaining the reduced efficiency.

A loss of dimerisation and functionality was noted in M5 constructs lacking an AT-ACP linker region. Comparison with other spirotetronate PKS modules identified similar regions in ChlA6 and KijS5, predicted to be remnants of KR domains. Creation of truncated constructs allowed the identification of an α-helix essential in module dimerisation and suggested a dimerisation motif was located towards the C-terminal of the linker. Cryo-EM of M5 showed a large dimer interface between the KS domains of two modules and a second contact region, likely to be the aforementioned dimerisation motif.

Abbreviations

ABC	ATP-Binding Casette
ACN	Acetonitrile
ACP	Acyl Carrier Protein
ADP	Adenosine Diphosphate
AEX	Anion Exchange
AMR	Antimicrobial Resistance
APS	Ammonium Persulfate
ARO	Aromatase
AT	Acyltransferase
ATP	Adenosine Triphosphate
bp	Base Pairs
bla	β-lactamase Gene
bom	Basis of Mobility
BSA	Bovine Serum Albumin
C	Condensation
CCD	Charge Coupled Device
CCK	Cholecystokinin
CD	Circular Dichroism
CDI	Carbonyldiimidazole
CID	Collision Induced Dissociation
CoA	Coenzyme A
Cryo-EM	Cryogenic-Electron Microscopy
CYC	Cyclise
Da	Daltons
DCM	Dichloromethane
dd	Docking Domain
DE	Dimerisation Element
DEBS	6-Deoxyerythronolide B Synthase

DH	Dehydratase
DIPEA	Diisopropylethylamine
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
DPCK	Dephospho-CoA Kinase
E. coli	*Escherichia coli*
EDC	1-Ethyl-3-(3-Dimethylaminopropyl) Carbodiimide
EDTA	Ethylenediaminetetraacetic Acid
EIC	Extracted Ion Chromatogram
EM	Electron Microscopy
ER	Enoyl-acyl Carrier Protein Reductase
ESI	Electrospray Ionisation
FAS	Fatty Acid Synthase
FDA	US Food and Drug Administration
FPLC	Fast Protein Liquid Chromatography
GC	Guanine-Cytosine
Gfp	Green Fluorescent Protein
GOI	Gene of Interest
HBS	HEPES Buffered Saline
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic Acid
His-tag	Polyhistidine-tag
HIV	Human Immunodeficiency Virus
HPLC	High Performance Liquid Chromatography
HSQC	Heteronuclear Single Quantum Coherence
HSV-1	Herpes Simplex Virus Type I
I	Ions (charge)
IMAC	Immobilised Metal Affinity Chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
K_a	On-rate
K_A	Kinetic Equilibrium Association Constant
K_d	Off-rate
K_D	Kinetic Equilibrium Dissociation Constant
KR	Ketoreductase
KR^0	Non-elongating Ketoreductase
KR^D	Dimerising Ketoreductase Domain
KS	Ketosynthase

KSAT	KS-AT didomain (with AT-ACP linker)
KSATS	KS-ATS didomain (without AT-ACP linker)
KS^0	Non-elongating Ketosynthase
lac	Lactose
LB	Lysogeny Broth
LC/MS	Liquid Chromatography Mass Spectrometry
LC-QTOF	Liquid Chromatography Quadruple Time-of-Flight
m/z	Mass Over Charge
LuxR	LuxR-type DNA-binding HTH domain
MALDI	Matrix-Assisted Laser Disorption Ionisation
MALDI-TOF	MALDI-Time-of-Flight
MRSA	Methicillin-resistant *Staphylococcus aureus*
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
MT	Methyltransferase
MW	Molecular Weight
MWCO	Molecular Weight Cut-off
Ni(II)	Nickel (II)
NADPH	Nicotinamide Adenine Dinucleotide Phosphate
NHS	N-Hydroxysuccinimide
NMR	Nuclear Magnetic Resonance
NOESY	Nuclear Overhauser Effect Spectroscopy
NRPS	Non-ribosomal Peptide Synthatase
NTA	Nitrilotriacetic Acid
PAGE	Polyacrylamide Gel Electrophoresis
PCP	Peptidyl Carrier Protein
PCR	Polymerase Chain Reaction
pEG	Polyethylene Glycol
PES	Polyethersulfone
pET151-D/TOPO	pET151 Directional TOPO
pH	Potential of Hydrogen
PKS	Polyketide Synthase
PMF	Peptide Mass Fingerprinting
Ppant	Phosphopantetheine
PPAT	Phosphopantetheine Adenylyltransferase
PPTase	Phosphopantetheinyl Transferase

PQQ	Pyrroloquinoline Quinone
PTM	Posttranslationally Modified
*p*TSA	*p*-Toluene Sulfonic Acid
QMN	Quartromicin
RBF	Rubidium Flouride
RBS	Ribosomal Binding Site
RI	Refractive Index
RF	Radiofrequency
RNA	Ribonucleic Acid
ROP	Repressor of Primer
SAM	*S*-Adenosyl Methionine
SARP	*Streptomyces* Antibiotic Regulatory Protein
SDM	Site-directed Mutagenesis
SDS	Sodium Dodecyl Sulfate
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
Sfp	Surfactin Phosphopantetheinyl Transferase
SNAC	*S-N* Acetyl-cysteamine
sp.	Species
SPR	Surface Plasmon Resonance
ssp.	Subspecies
TB	Tuberculosis
TBE	Tris-boric Acid EDTA Buffer
TE	Thioesterase
TEAA	Triethylammonium Acetate
TEM	Transmission Electron Microscope
TEV	Tobacco Etch Virus
TFA	Trifluoroacetic Acid
THF	Tetrahydrofuran
TLC	Thin Layer Chromatography
TOF	Time-of-Flight
Tris-HCl	Trisaminomethane-hydrochloride
[U-^{15}N]	Uniformly Labelled With ^{15}N Isotopes
UV	Ultra Violet
WHO	World Health Organisation

Chapter 1

Introduction

1.1 Natural products

1.1.1 The history of natural products

Natural products have been used as therapeutic remedies for millennia. Indigenous cultures were able to utilise the remedial properties of many organisms to create the first crude drugs, although these shamanistic cultures attributed the health benefits to magic. This primitive belief was held until civilisations began to develop a deeper scientific understanding of medicine. The Huangdi Neijing, an ancient Chinese text likely originating from before the Han dynasty ($\sim$ 210 BC), was one of the earliest books to link health and the progress of disease to the environment, poor diet and hereditary factors.[1] In addition, this text also refers to numerous natural remedies and their benefits in preventing and curing disease. Thenceforth, the development of pharmaceuticals has been closely linked to pharmacognosy. This can be seen in Arabic, Greek and Roman medical texts such as *De Materia Medica* (1 BC), which documents over 600 medicinal plants and their applications.[2] The dissemination of knowledge throughout Europe led to further developments by apothecaries, such as Nicholas Culpeper (1616 - 1654), and the establishment of guilds with the purpose of cultivating plants for medicinal purposes.

It was not until 1817 that an active principle was purified from a natural source when Friedrich Sertürner isolated morphine from *Papaver somniferum* (opium poppy).[3] Later, in 1828, Johann Buchner isolated salicin from willow bark,[4] which subsequently led to Bayer deriving the analogous drug, aspirin.[5] The first antibacterial natural product was later discovered when Alexander Fleming accidentally developed mould on plates containing *Staphylococcus aureus* and noticed that the colonies around the mould had been destroyed.[6] Penicillin was subsequently isolated from the mould and shown to be the active principle with activity against *Staphylococcus aureus*. The following period resulted in a golden age of drug discovery,[7] with many natural products being identified and developed into drugs (Figure 1.1). These natural products and natural product derivatives include a wide range of antibacterials (colistin, ampicillin, vancomycin, rifamycin),[8–11] anticancer agents (vincristine, doxorubicin)[12;13] analgesics (morphine, oxycodone)[14] and antimalarials (quinine, proguanil).[15]

Despite the rise in the number of synthetic drugs entering the market, natural products

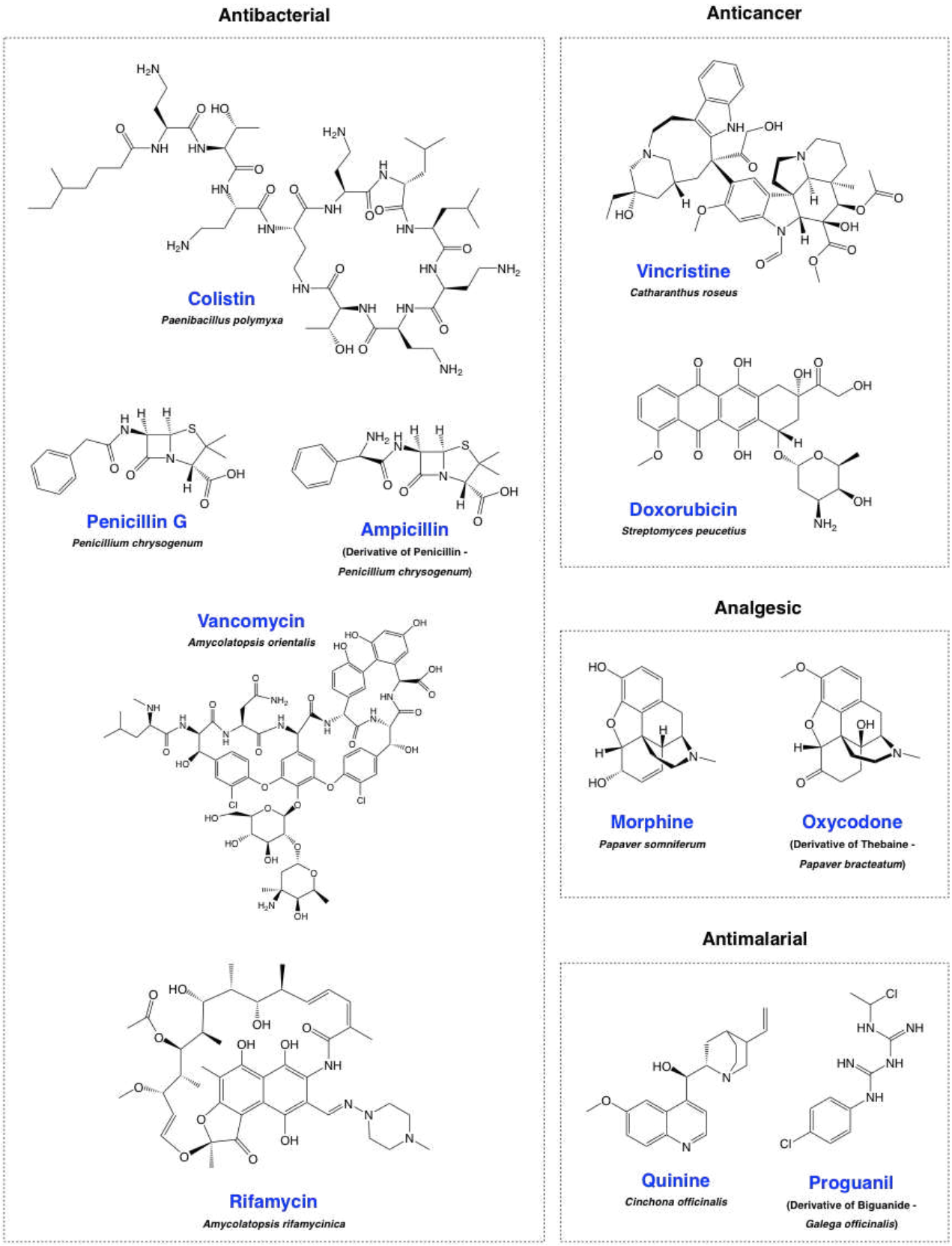


Figure 1.1. *Examples of natural products and natural product derivatives with antibacterial, anticancer, analgesic and antimalarial properties.*

have remained an important source of novel drug discovery.[16] 34% of medicines approved by the US Food and Drug Administration (FDA) between 1981 and 2010 were natural products or natural product derivatives.[17] Furthermore, with less than 1% of the microbial community estimated to have been cultured in a lab,[18] natural products will remain an abundant source of pharmaceuticals for years to come.

1.1.2 Antimicrobial resistance

Although a large number of antimicrobials are available, antimicrobial resistance (AMR) is an increasing problem. It is now thought that over 70% of pathogenic bacteria are resistant to at least one antibiotic.[19] With globalisation increasing the spread of diseases, it is now more imperative than ever to produce novel drugs. This has lead to the World Health Organisation (WHO) categorising AMR as a serious threat to global public health and stating that it requires action across all sectors of government and society.[20]

For example, colistin (also known as polymyxin), discovered in the 1940s,[8] was rarely used due to a number of negative side-effects and remained an effective last resort antibiotic. However, over the past few years, numerous cases of genes that confer colistin resistance through horizontal gene transfer have been found, including cases in the UK and US.[21]

1.2 Polyketide synthases

Many pathways exist that contribute towards the production of natural products, including non-ribosomal peptide synthetase (NRPS),[22–25] fatty acid synthase (FAS)[26;27] and polyketide synthase (PKS).[28] The structurally diverse PKSs are a family of multi-enzyme biosynthetic assembly lines that produce secondary metabolites, known as polyketides.[29]

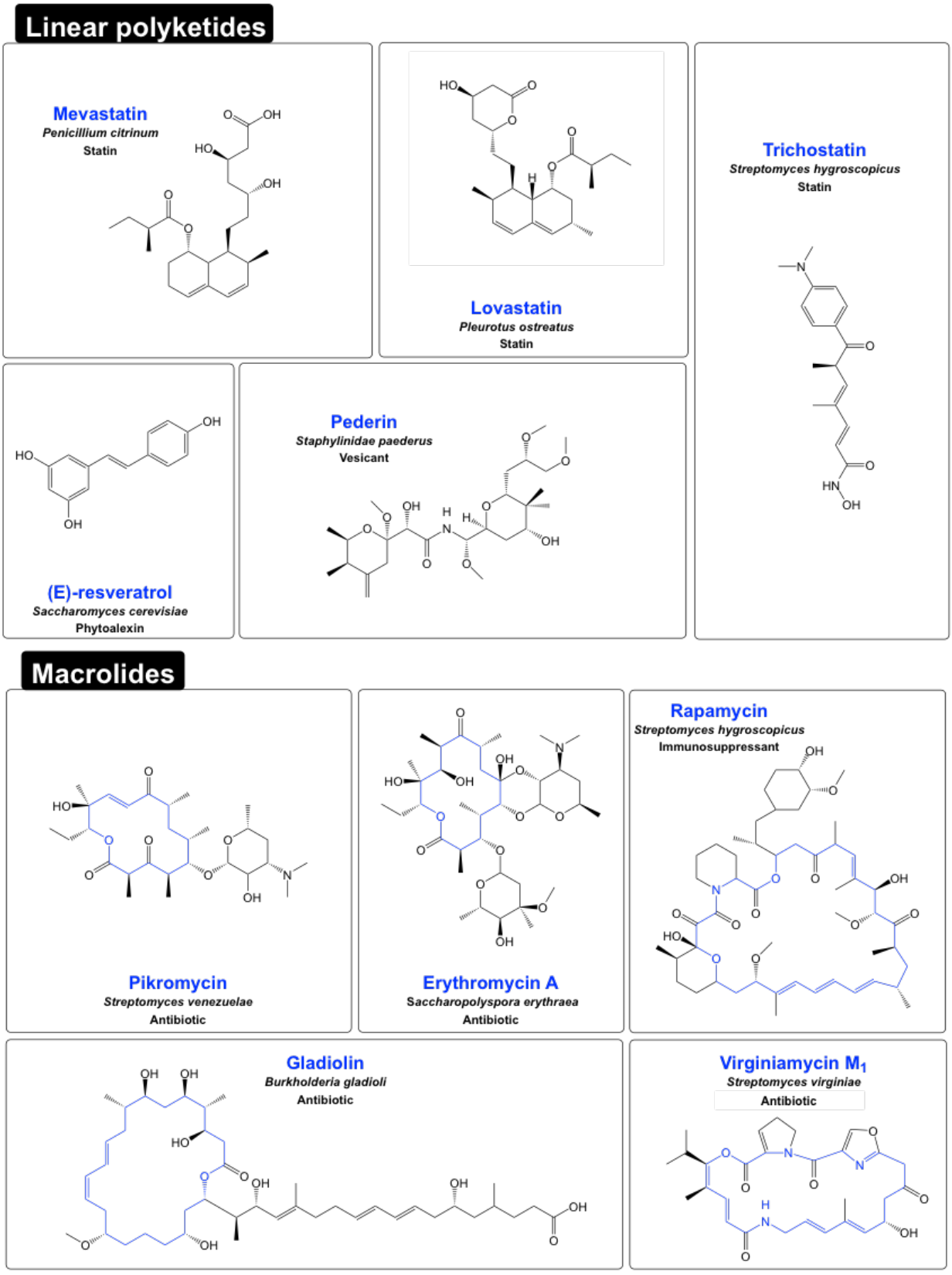


Figure 1.2. *Examples of natural products produced by polyketide synthase pathways. Linear polyketides (top) differ from macrolides (bottom) as the polyketide does not form a macrocycle.*

1.2.1 Polyketides

Polyketides form one of the largest groups of natural products and have a very diverse range of biological activities (Figure 1.2).[30–32] The first polyketides to be discovered were linear, such as lovastatin (lowers cholesterol)[33] and resveratrol (dietary supplement).[34] In 1949, a second class was discovered[35] when pikromycin was isolated from *Streptomyces venezuelae*.[36] Pikromycin was categorised as a macrolide, as it contains a macrocyclic lactone ring. Since pikromycin's discovery, numerous other macrolides with varying activities have been isolated, such as gladiolin (activity against *Mycobacterium tuberculosis*)[37] and erythromycin (activity against *Staphylococcus aureus* and *Haemophilus influenzae*).[38–40]

Despite their expansive structural diversity and range of biological activity, polyketides are all derived primarily from acyl and (methyl)malonyl thioester building blocks that undergo a series of decarboxylative Claisen condensations catalysed by a polyketide synthase (PKS).

1.2.2 Polyketide synthase classification

Functional differences have led to PKSs being separated into three distinct classes: types I, II and III.[41] Type I PKS are large multifunctional protein complexes that are organised into modules.[42] Each module contains the enzymatic machinery required for one elongation step, as well as any additional domains such as a ketoreductase (KR),[43] dehydratase (DH) or an enoyl-acyl carrier protein reductase (ER).[44;45] Further subdivision of type I PKSs will be discussed in section 1.2.4. Type II PKS consist of an aggregate of monofunctional proteins that adopt an iterative mechanism.[46] Minimal type II PKS modules are comprised of two ketosynthase (KS) domains and an acyl carrier protein (ACP), with further optional reductive domains, cyclases (CYC) and aromatases (ARO) present in non-minimal modules.[47] Type III PKS contains a single protein able to catalyse a series of condensation reactions directly from an acyl-CoA substrate to generate a polyketide.[30;48] Although the vast majority of type III PKSs are found in plants, examples have also been found in bacteria and fungi.[49]

1.2.3 Polyketide synthase domains

ACPs are the central component of both type I and type II PKS modules and contain four conserved helices that display a high degree of conformational flexibility.[50] They play a vital role in the translocation of polyketide chains between modules and are known to interact with over thirty different enzymes.[51;52] The first ACP was structurally characterised in 1988,[53] showing the first long conserved helix to run antiparallel to helix 2 and 3. Further solved structures showed a fourth shorter helix is usually present with varying orientations (Figure 1.3).[54] The second helix and the shorter fourth helix form a hydrophobic pocket. A conformational equilibrium was noted for a number of ACPs[55], with the first conformation seeming to allow the polyetide chain to residue in the hydrophobic pocket to protect the polyketide chain from hydrolysis. The second conformation allows the polyketide chain to withdraw from the hydrophobic pocket and to be processed by partern enzymes.[56;57]

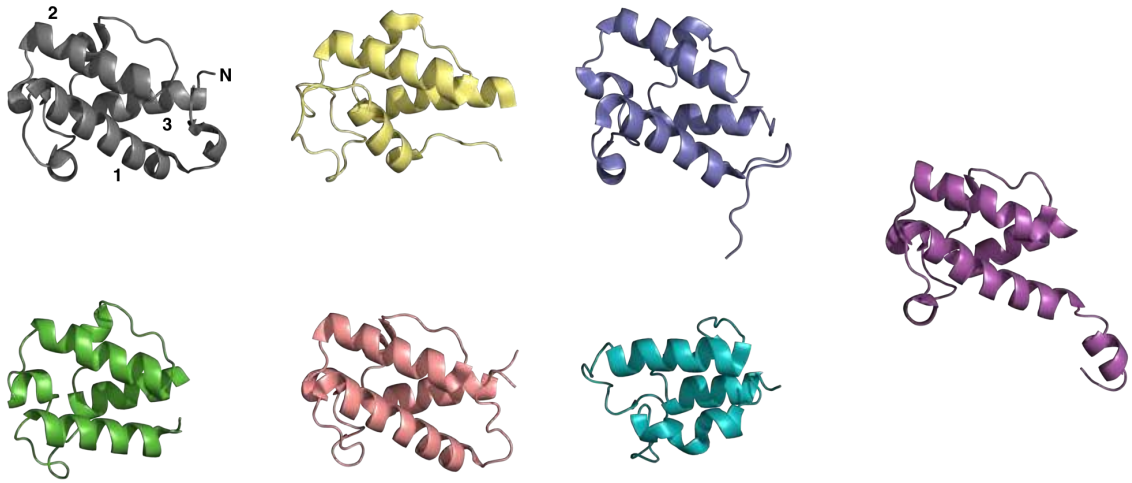


Figure 1.3. *Solution NMR structures of ACPs from modular type I PKSs (PDB: 5hv8, 2mvu, 2n98, 2n50, 5hvc, 1acp and 2ju2). Helices 1, 2 and 3 are labelled.*

In nature, each ACP is in its *holo*-form,[58] meaning a phosphopantetheine arm (Ppant) is attached to the substrate attachment site (serine residue, Figure 1.4 (a)).[59] The Ppant arm is approximately 20 Å in length, allowing the ACP to interact with the module's other domains, before transferring the nascent chain to a downstream module.[60–62]

Type I PKS pathways commonly start with a loading module containing an acyltransferase (AT) and ACP domain.[63] A KS domain that does not sequester a polyketide, due to the replacement of a conserved active site cysteine with a glutamine, may also be present

(KS^Q).[64] AT domains ($\sim$ 420 residues) consist of two subdomains: a large hydrolase core, and a smaller ferredoxin-like subdomain. A catalytic serine residue, located in the channel between the two subdomains, attacks the thioester bond of a CoA-bound starter unit before the terminal thiol group of the adjacent ACP's Ppant arm attacks the resulting ester bond. This thioester exchange reaction results in the transacylation of a starter unit onto the ACP's Ppant appendage (Figure 1.4(b)).[65]

Downstream of the loading module are a series of modules that extend and modify the starter unit to produce a polyketide. The minimal set of enzymes required for a type I PKS module to extend a polyketide are the KS, AT and ACP domains, which are collectively responsible for chain elongation.[66] The AT catalyses transacylation of an extender unit, usually (methyl)malonyl, onto the terminal thiol of the ACP's Ppant appendage using the same mechanism as in the loading module (Figure 1.4(c)).[67] KS domains ($\sim$ 430 residues) contain a conserved thiolase fold ($2\alpha5\beta2\alpha5\beta2\alpha$) that forms the domains central core.[68–70] The thiol of the KS's catalytic cysteine residue attacks the thioester of the upstream ACP-bound intermediate, resulting in a new thioester bond to the KS (Figure 1.4(d)). The KS bound polyketide and the (methyl)malonyl extender unit tethered to the intramodular ACP then undergo a decarboxylative Claisen condensation, resulting in the transfer of the polyketide from the KS to the intramodular ACP and an extension of the chain by two carbons.

The additional presence of optional reductive domains in each module adds diversity to the produced polyketides. KR domains ($\sim$ 440 residues) contain two subdomains, both similar to a short-chain dehydrogenase/ reductase.[71] One subdomain contains a truncated Rossmann fold and only plays a role in maintaining the structural integrity of the KR. The second subdomain catalyses the stereoselective NADPH-dependent reduction of the β-keto group to a β-hydroxy (Figure 1.4(e)). KRs can be categorised into three distinct subclasses. A-type KRs, containing a conserved tryptophan residue, result in the production of a L-3-hydroxyacyl intermediate and B-type KRs, containing a conserved LDD motif, result in the production of a D-3-hydroxyacyl intermediate. It is proposed that the different stereochemistry is achieved by the A and B-type KRs guiding the substrates to the active site from opposite directions.[72;73] C-type KRs are catalytically inactive.

DH domains ($\sim$ 280 residues) contain a double hotdog fold ($3\beta\alpha4\beta\alpha3\beta\alpha4\beta$)[74] that forms a tunnel that the substrate must enter to access the active site.[75] The active site contains

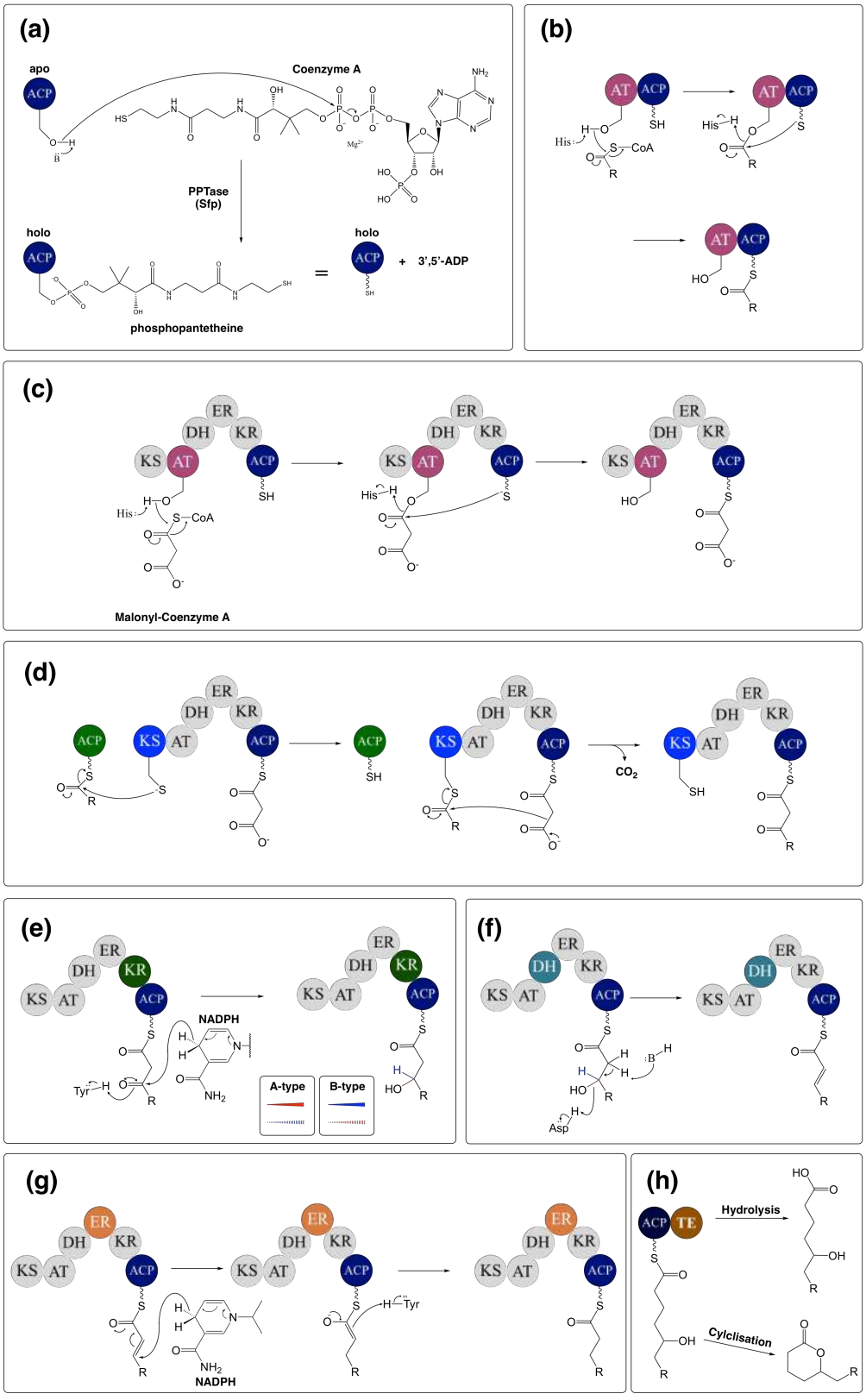


Figure 1.4. ***(a)*** *The conversion of apo-ACP to holo-ACP,* ***(b)*** *the mechanism of a PKS loading module and the function and mechanism of PKS domains;* ***(c)*** *acyltransferase (AT),* ***(d)*** *ketosynthase (KS),* ***(e)*** *ketoreductase (KR),* ***(f)*** *dehydratase (DH),* ***(g)*** *enoyl reductase (ER) and* ***(h)*** *thioesterase (TE).*

conserved histidine and aspartic acid residues that catalyse dehydration of the hydroxyacyl intermediate to produce an α,β-unsaturated acyl intermediate (Figure 1.4(f)).

The presence of an ER domain ($\sim$ 310 residues), containing both catalytic and structural subdomains, can catalyse the NADPH-dependent reduction of the C=C bond (Figure 1.4(g)). This reaction is hypothesised to be reliant on a proton being provided by a tyrosine residue in the ER active site.[76]

Other domains may also be present, including a methyltransferase (MT). MTs utilise *S*-adenosyl methionine (SAM) to transfer a methyl group onto the α-position of the polyketide (Figure 1.5).[77]

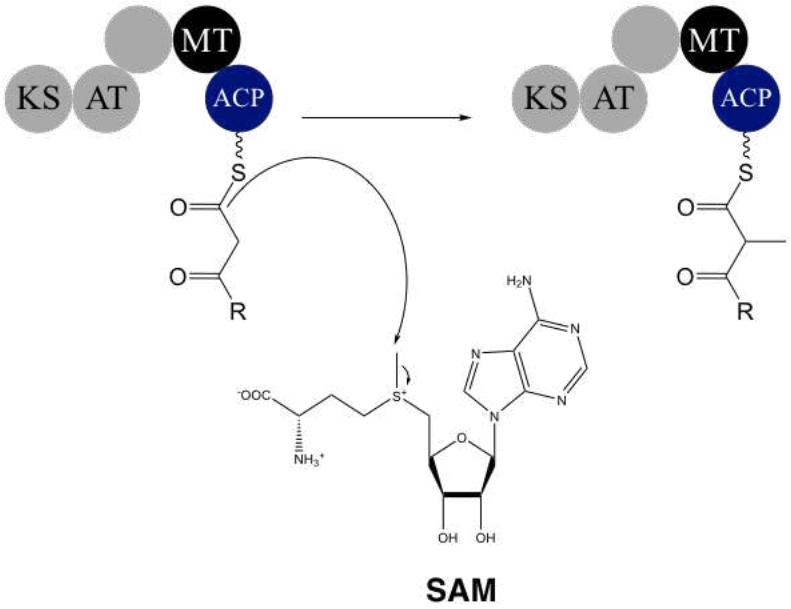


Figure 1.5. *The mechanism of an MT domain loading a methyl group to the α-position of a polyketide chain.*

Once all rounds of chain extension have been completed, cleavage of the natural product from the PKS enzymes occurs. This is commonly catalysed by a thioesterase (TE, $\sim$ 270 residues) that is present at the C-terminal of the final PKS module, either through hydrolysis or cyclisation (Figure 1.4(h)). The released chains are often precursors to the final natural product, with post-PKS tailoring enzymes catalysing further modifications, such as hydroxylations, glycosylations and methylations.[78] In addition, acyltransferase-like enzymes fulfill housekeeping/proofreading roles by removing abherrant polyketides from the PKS pathway through hydrolysis.[79]

1.2.4 Subdivisions within type I polyketide synthases

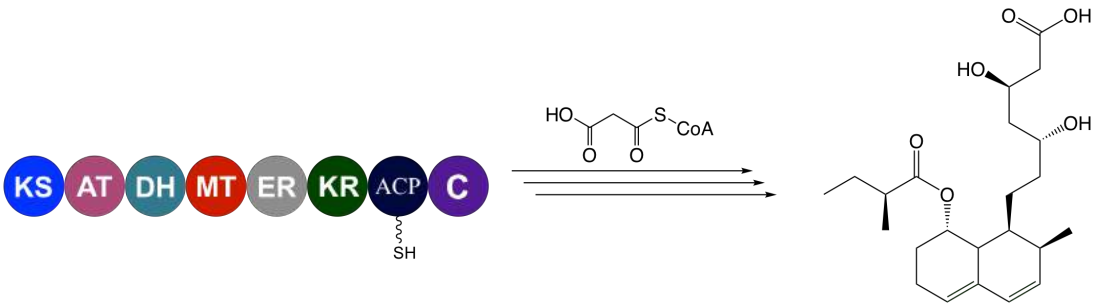


Figure 1.6. *Mevastatin is produced by an iterative PKS that uses malonyl-CoA extender units. In each catalytic cycle, different levels of reduction occur depending on which domains act on the nascent chain.*

Type I PKS systems may either be categorised as iterative or modular.[80] Iterative PKS systems contain a module with the ability to catalyse repetitive elongations and modifications. In each catalytic cycle, different domains modify the polyketide chain, resulting in varying levels of reduction.[81] One example of an iterative PKS is the enzyme in mevastatin production (Figure 1.6), in which the KS, AT and KR domains are utilised in each of the eight catalytic cycles. The DH domain is employed in all but two cycles, resulting in the presence of two hydroxyl groups, whilst the ER domain is only utilised four times, resulting in the presence of C=C bonds[82] that undergo a Diels Alder reaction to form the cyclised core.[83] In addition, the MT domain creates the methyl-branch of the α-methylbutyryl ester moiety. A condensation (C) domain, commonly found in non-ribosomal peptide synthetase (NRPS),[22] is also present and likely plays a role in polyketide chain release.

In contrast, modular PKSs utilise numerous modules to extend and modify a polyketide product, with each module typically catalysing one round of chain extensions.[85] The archetypal example of a modular PKS is 6-deoxyerythronolide synthase (DEBS),[86–89] which utilises six methylmalonyl and one propionyl building block to produce a precursor to erythromycin A (Figure 1.7).[84;90]

Modular type I PKSs may be further categorised into either *cis*- or *trans*-AT systems. In *cis*-AT PKS, such as DEBS PKS, every module contains an AT domain, whereas *trans*-AT PKS's modules use stand-alone AT domains which service multiple modules.[91;92] Hybrid PKS systems also exist, such as enacyloxin, with some modules containing an AT domain and others employing a stand-alone AT.[93;94]

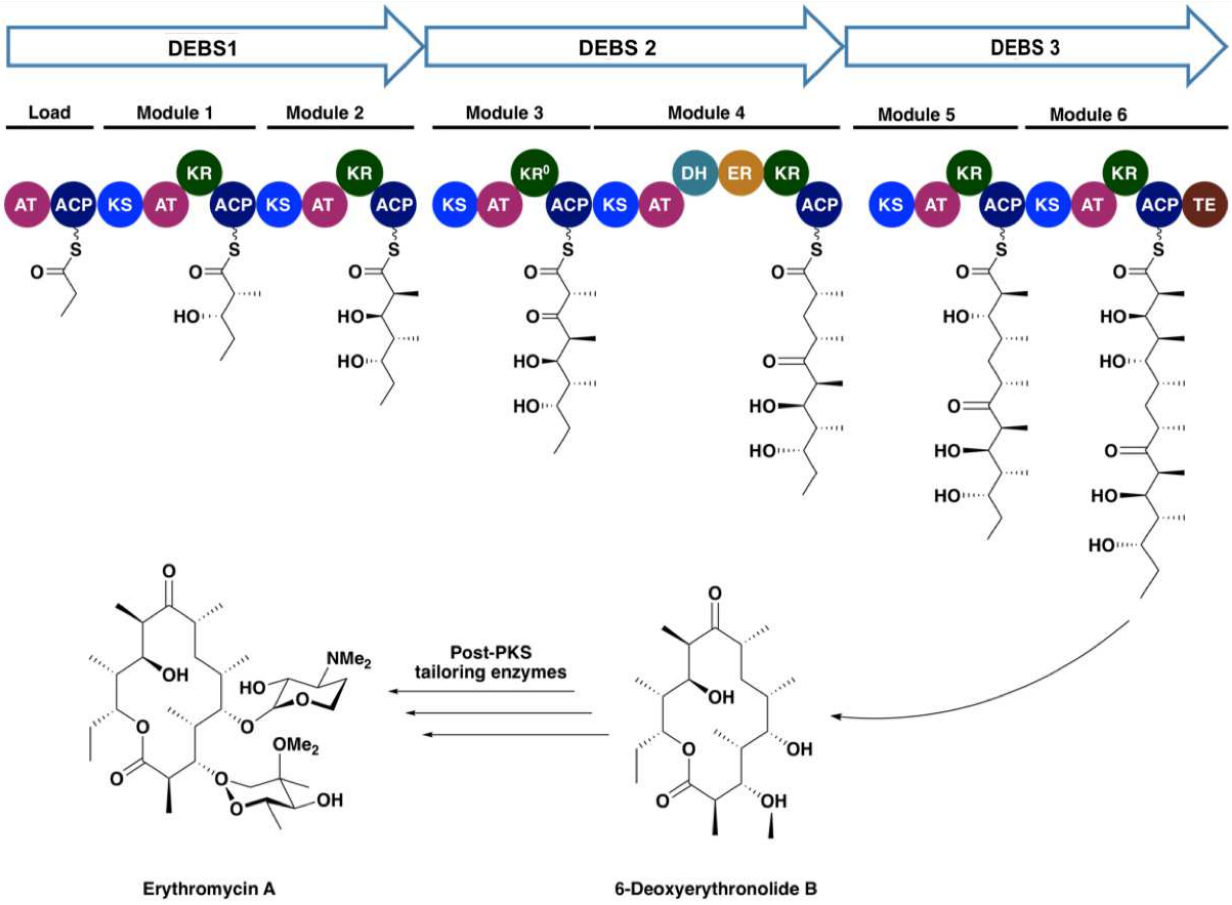


Figure 1.7. *The DEBS PKS pathway and the post-PKS tailoring enzymes that act upon the resulting precursor, resulting in the production of erythromycin A.*[84]

1.2.5 Discovering polyketide synthases and product prediction

Type I modular PKSs generally follow the colinearity rule, meaning the domains are utilised in the order they are encoded in the genome.[95] Since sequence similarity is highly conserved amongst PKS domains, their activities including extender unit selection,[96] sterochemistry[97] and catalytic functionality,[98] can often be predicted by sequence comparisons. This makes it possible to identify the PKS likely responsible for the production of a particular polyketide metabolite by analysis of genomic DNA.[99]

Although this should mean the product of a PKS can be predicted based on sequence analysis, in reality, the polyketide product is not necessarily dictated by the order of the PKS encoding genes in the cluster. This is the case in calcimycin PKS, where the final module is encoded by a gene located before the gene that encodes the two upstream modules in the cluster.[100] In addition, domain/module skipping and stuttering

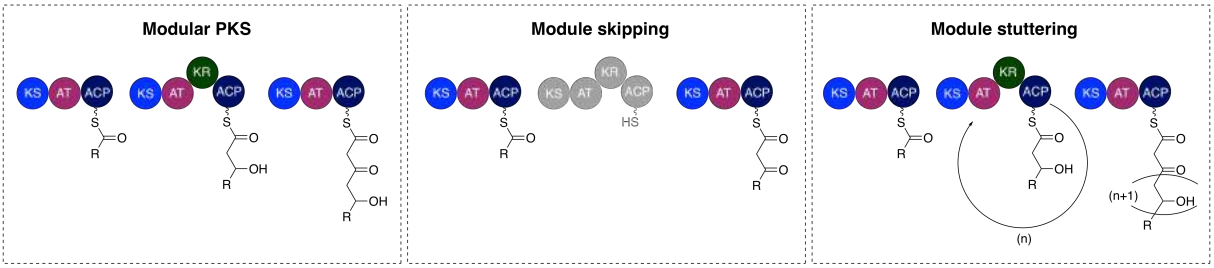


Figure 1.8. *Schematic depiction of possible modular PKS module skipping and stuttering mechanisms.*

provide interesting scenarios where modules may either be bypassed or catalytically modify the polyketide multiple times (Figure 1.8).[101;102] An example of module skipping is in pikromycin/methymycin PKS from *Streptomyces venezuelae*,[103] where two different ACPs can interact directly with a TE. This creates the possibility of chain termination occurring before the final modification and extension is catalysed by the stand-alone final module (Figure 1.15).[104]

1.2.6 Dimerisation interfaces for polyketide synthase subunits

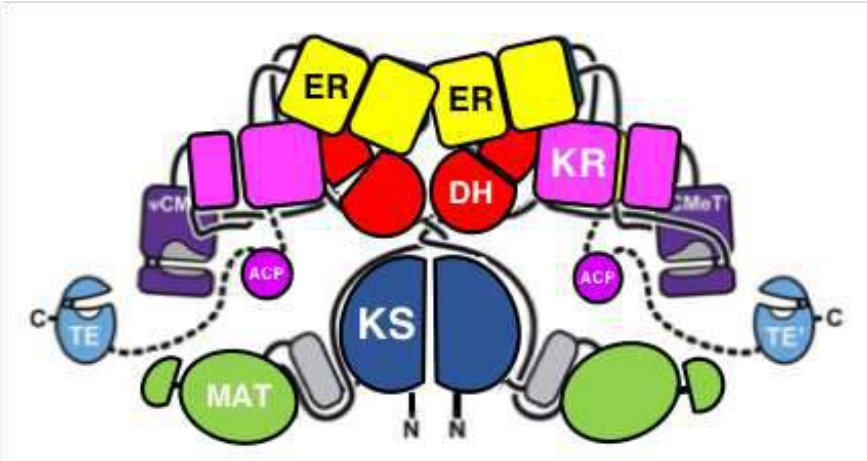


Figure 1.9. *Schematic representation of porcine FAS*

Previously, the understanding of type I PKSs was based upon the organisation of FAS modules, which shares many of the same domains as PKS due to their intertwined evolutionary history.[105;26;106] FAS may either be type I, common in fungi and mammalian cells, or type II, common in archaea and bacteria. Type I FAS consist of large multifunctional polypeptides, whereas type II FAS use discrete monofunction enzymes.[107] Structural elucidation of FAS domains through X-ray crystallography was successful for all common

domains, including a 4.5 Å structure of a full-length FAS (Figure 1.9). The solving of crystal structures of a KS-AT didomain, DH, ER-KR and a TE, as well as an NMR structure of an ACP, provided a vital insight into the boundaries of domains within PKSs, as well as highlighting common differences between the domain organisation of PKS and FAS systems. Further acquisition of domain/didomain structures showed AT, KR, ER and ACP domains to be commonly monomeric, allowing the flexibility for the conformational changes required for PKS functionality (Figure 1.10).[33;71;76;108–117]

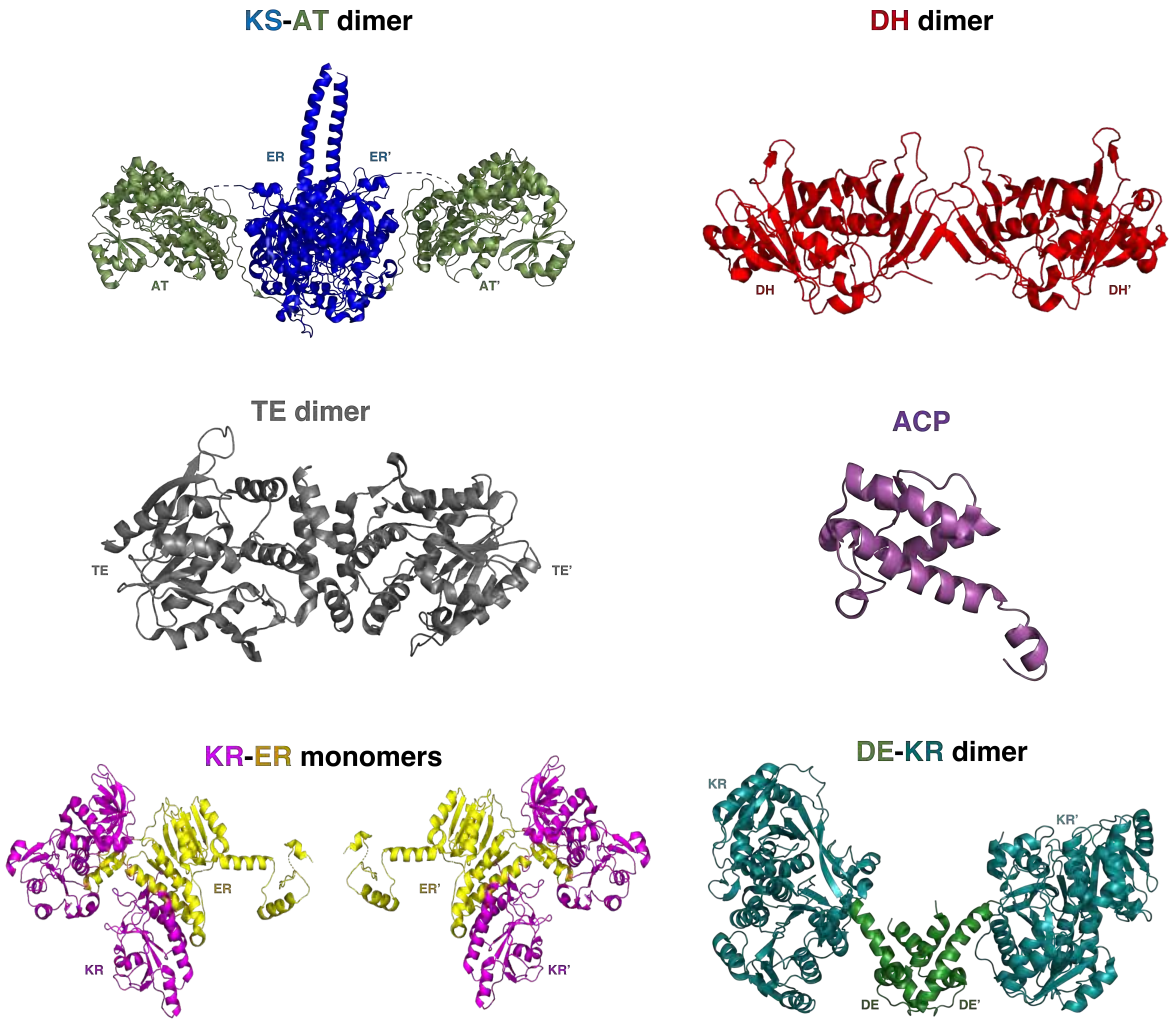


Figure 1.10. *Crystal structures of a KS-AT didomain (PDB: 2HG4), DH (PDB: 3KG7), KR-ER (PDB: 3SLK) didomain, TE (PDB: 1KEZ) domain and a DE-KR complex (PDB: 4IMP) , and an NMR structure of an ACP (PDB: 2JU2).*[118]

In archetypal type I PKS systems, the dimerisation of modules is predominantly due to the KS, which often has a large dimeric interface (Figure 1.11).[74] The only other domains that commonly make contact intramodularly are the DH and TE domains, both of which have a much smaller interface than the KS domains.[75] Each dimerising surface contains highly conserved accessory motifs.

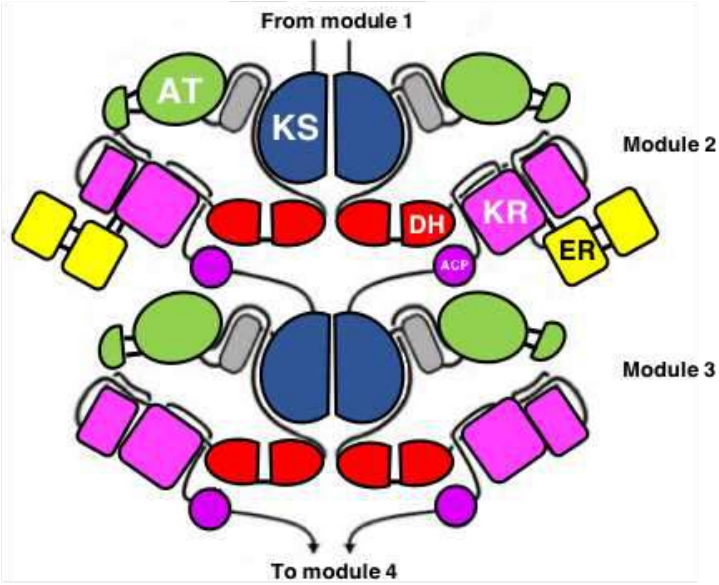


Figure 1.11. *A schematic representation of a type I multimodular PKS (spinosyn).*[76;119] *Linkers are shown in grey.*[120]

An additional dimerisation element (DE) exists in approximately 50% of β-modules.[121] Located between the C-terminus of an AT and the N-terminus of a KR, these highly conserved $\sim$ 55 residue DEs have been shown to facilitate modular dimerisation.[122]

1.2.7 The overall architecture of polyketide megasynthases

Although structures of PKS regions were obtainable through NMR and X-ray crystallography, the limitations of both techniques prevent their use in the acquisition of structural data for an intact module. However, recent advances in cryo-EM technology led to the first structural insights into the catalytic rearrangement of an intact PKS module, when models of the penultimate module of pikromycin PKS (PikAIII, β-module) were acquired for each conformation of its catalytic cycle (Figure 1.12).[117;123] The dimeric *apo*-PikAIII was observed to be in two varying conformations, both of which showed the KS, AT and KR domains form a chamber. In both conformations, the ACP was located inside the chamber and either interacting with the KR domain or in close proximity to the KS and AT domains.

The structural elucidation of varying constructs allowed a greater insight into the intramodular enzymatic mechanisms and interactions with the upstream ACP. Firstly, a truncated construct, omitting the ACP (ACP5), was structurally elucidated (PikAIII/-

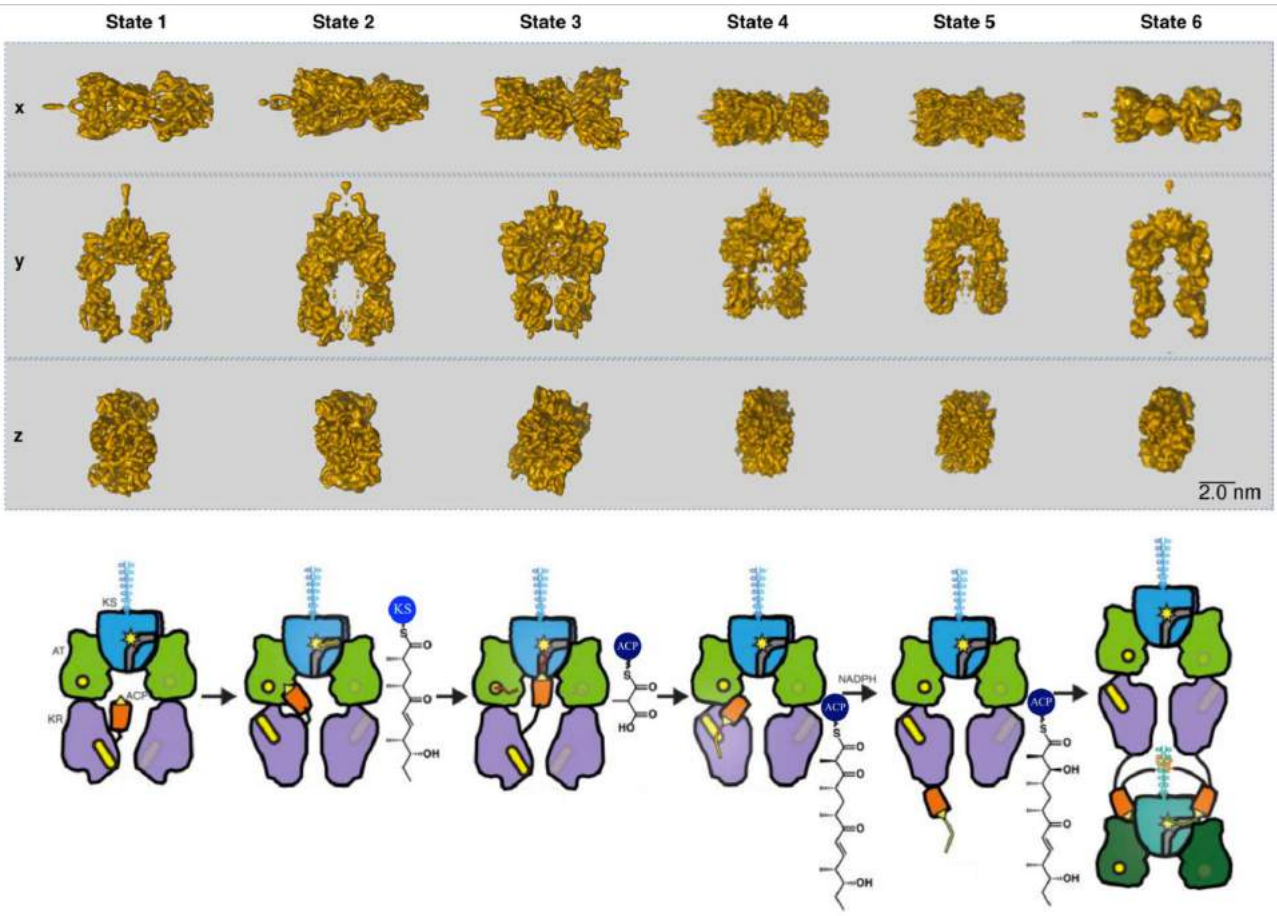


Figure 1.12. *Cryo-EM models (top) and corresponding schematic representation of substrate processing in PikAIII (bottom). A pentaketide substrate is transferred from an upstream ACP to the KS domain of holo-PikAIII (**state 1**), inducing a rotation of the KR domain and relocation of the intramodular ACP domain to the AT in preparation for loading with an extender unit (**state 2**). Loading holo-PikAIII with a methylmalonyl-extender unit results in the repositioning of the ACP to the KS's bottom entrance (**state 3**). Incubation with a β-ketohexaketide results in the relocation of the ACP to within close proximity of the KR active site in preparation for reduction of the polyketide (**state 4**) and addition of NADPH leads to an ejection of the ACP out of the central chamber as a result of a KR-catalysed reduction of the polyketide to a β-hydroxyhexaketide (**state 5**). This places the ACP in a position for intermediate transfer to PikAIV (**state 6**).*

ΔACP5). The removal of the ACP resulted in a rotation of the KR domains around the AT domains by approximately 165°, suggesting the intramodular ACP plays a role in the positioning of the KR relative to the AT domain.

Secondly, the intermodular interaction with the upstream ACP of PikAII (ACP4) was explored. A fused ACP4-PikAIII/ΔACP5 construct was created and the ACP was loaded with a pentaketide to encourage interactions with PikAIII's KS domain. The resulting structure showed that ACP4 binds to the KS domain outside PikAIII's chamber. The *holo*-form of a KS active-site mutated variant (*holo*-ACP4-PikAIII/C209A/ΔACP5) was used as a control to confirm that the location of ACP4 was due to the conformational changes required to facilitate pentaketide translocation.

Thirdly, the intramodular machinery of PikAIII was studied. Loading of a methylmalonyl extender unit onto PikAIII resulted in the ACP relocating to a second interaction site on the KS, this time inside the PikAIII chamber. This provided evidence that the KS has separate entrances for substrate loading from the upstream ACP (side entrance) and subsequent offloading to the intramodular ACP (bottom entrance).

Incubation of PikAIII with a pentaketide resulted in β-ketohexaketide-PikAIII and the addition of NADPH yielded β-hydroxyhexaketide–PikAIII. Structural elucidation showed that, in both cases, the KR and AT domains were in close proximity and the AT domain occluded the KS's side-entrance, thus preventing substrate transfer from ACP4 to the KS of PikAIII. In the case of β-ketohexaketide-PikAIII, ACP5 is docked with the KR's lid helix, effectively presenting the β-ketohexaketide to the KR's active site and allowing reduction to β-hydroxyhexaketide in the presence of NADPH. Upon addition of NADPH, the ACP is released and relocated outside the central PikAIII chamber, facilitating further interactions with PikAIV.

1.2.8 Docking domains

PKSs require selective intersubunit interactions to ensure the correct biosynthetic route is maintained for the consistent assembly of products. Interactions between subunits have been shown to be promoted by docking domains located at the C-terminus of an upstream subunit (ACPdd) and the N-terminus of a downstream subunit (KSdd).[124] The ACPdds are connected to the ACP's four conserved helices via a linker region (30 - 50 residues) that provides a degree of flexibility; whereas KSdds are located directly before the downstream module's KS domain. These domains promote intermodular crosstalk and facilitate the transfer of polyketide intermediates from an upstream ACP to a downstream KS. Moreover, the specificity of docking domain interactions plays a vital role in maintaining the correct order of modules in the PKS biosynthetic pathway, ensuring the consistent production of the required polyketide.

The structural elucidation of ACP constructs containing docking domains has previously been unsuccessful, perhaps due to the mobility of the linker region. Furthermore, KSdds are often insoluble or overexpress in low quantities when separated from the rest of the KS domain. However, by creating fused constructs containing the upstream ACPdd

and downstream KSdd, soluble proteins have been produced.[124] This allowed docking domain complexes to be structurally elucidated from the biosynthetic pathways of pikromycin[125], curacin,[126;127], erythromycin,[124] and virginiamycin.[128] The structural organisation varies in each case (Figure 1.13). Characterised structures, coupled with sequence analysis of identified docking domains has led to *cis*-AT PKS docking domains being characterised into three different phylogenetic classes that have been shown to be mutually incompatible.[129]

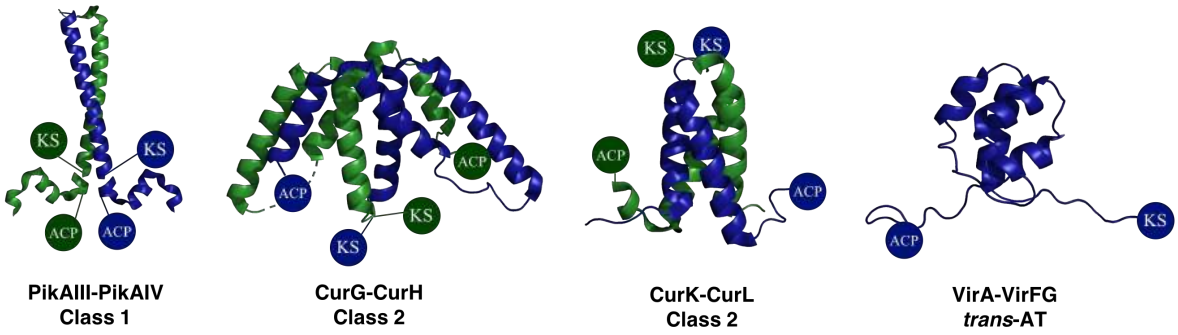


Figure 1.13. *Solved structures of interacting docking domains from PKS systems: CurG-CurH (PDB: 4MYY), CurK-CurL (PDB: 4MYZ), PikAIII-PikAIV (PDB: 3F5H) and VirA-VirFG (PDB: 2N5D).*

Existing structures of fused class 1 docking domains, found in actinobacteria PKSs,[126] show each ACPdd consists of two dimerisation helices that form a 4-helix bundle with a second ACPdd[130] and a C-terminal docking helix that plays a vital role in facilitating interactions with a downstream KSdd (Figure 1.14). Class 1 KSdds consist of a long α-helix that dimerises with a second KSdd to form a coiled-coil. The third helices of the ACPdd dimer then bind to each side of the coiled-coil, interacting with opposite KSdds of the dimer pair. The specificity of intermodular protein-protein interactions is thought to be promoted by the location of electrostatically charged residues on the available interface, with interactions between the ACPdd's C-terminal helix and the KSdd thought to be of particular importance.[131;132] Interestingly, the upstream ACPs are initially directed away from the downstream KS domains.[126] The presence of a linker region between each ACP and ACPdd is therefore vital in providing the required mobility for the ACP to relocate to the KS after intermodular docking has occurred, allowing substrate transfer.

Class 1 docking domains may be further categorised into three distinct subclasses based upon highly conserved structural motifs. Although domains from different subclasses sometimes exhibit interactions, they display a significantly weaker affinity than compatible docking domains.[125]

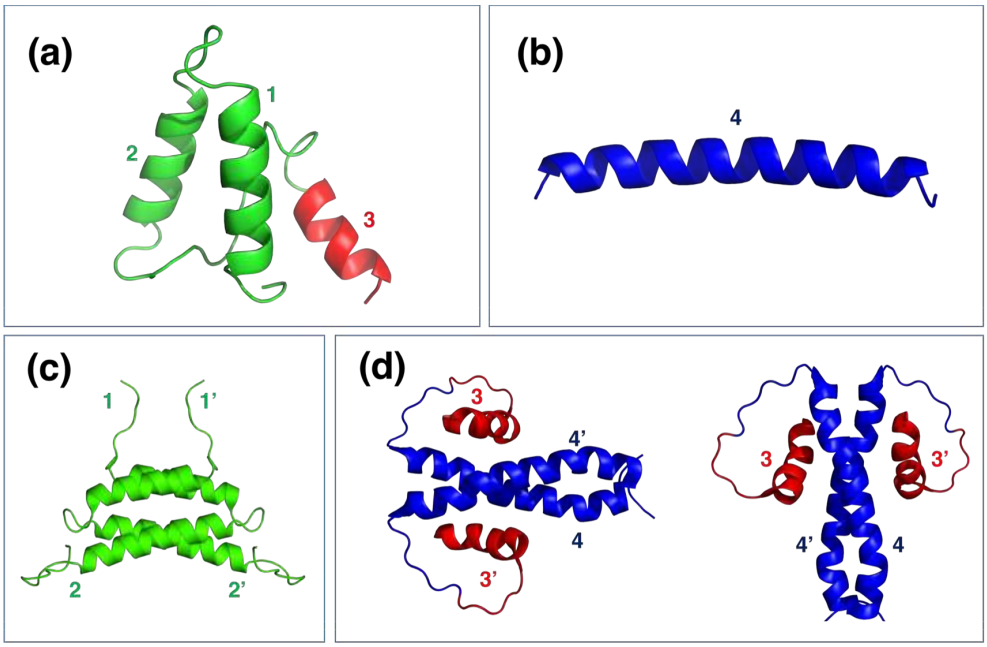


Figure 1.14. *Homology models of a class 1 ACP docking domain (**a**, DEBS2) and a class 1 KS docking domain (**b**, DEBS3). NMR structures of a fused construct showed dimerisation occurs between helices 1 and 2 of the ACPdd (**c**, PDB: 1PZQ) and helix 3 interacts with the coiled-coil formed by two KSdds (**d**, PDB: 1PZR).*

Class 2 docking domains are found in cyano- and myxobacteria PKSs. Crystal structures for two pairs of class 2 docking domains from curacin PKS showed that, unlike the class 1 docking domains, both the KSdd and ACPdds consisted of two α-helices connected by a sharp bend.[126] The ACPdd was found to be approximately forty residues shorter than for class 1 ACPdd and the KSdd to be a similar length. Formation of a coiled-coil between the second helix of a pair of KSdd's was observed and the first helix of each docking domain was seen to extend away from the dimer interface due to the lack of hydrophobic surface to support the formation of a coiled-coil. The second helix of each ACPdd then binds to the KSdd's coiled-coil, interacting with both KSdd's in the coiled-coil. This is in direct contrast to class 1 docking domains, where a pair of ACPdd's third helices bind to either side of the KSdd's coiled-coil. This leads to the ACP being directed towards the KS domain, resulting in more efficient promotion of acyl-chain transfer between the subunits.

Class 3 docking domains are found in PKS-NRPS hybrids,[133] such as tubulysin, and facilitate interactions between peptidyl carrier proteins (PCP) and the C domain of a compatible module.[128] The N-terminal docking domain of the C domain has been shown to adopt an $\alpha\beta\beta\alpha\alpha$ topology, allowing dimerisation to occur between two C domains.[134]

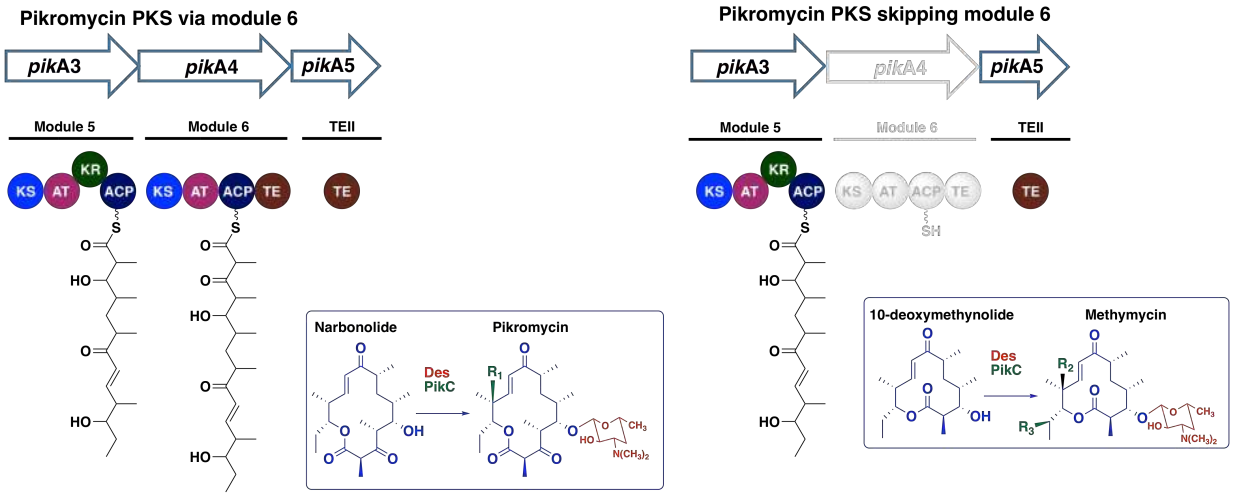


Figure 1.15. *The proposed PKS pathways for pikromycin/methymycin PKS[36] production and the resulting products.*

Docking domains have been shown to retain a high degree of selectivity once disconnected from their parent subunit.[125] A number of discrete docking domain constructs were produced and interactions were measured by surface plasmon resonance (SPR), resulting in binding only occurring between compatible pairs. In addition, the fusing of compatible docking domain pairs onto different parent subunits has been shown to facilitate intersubunit acyl transfer, even if the class of the replacement docking domain does not match the class of the original docking domain.[126] Pikromycin PKS produces both narbonolide (nbl) and 10-deoxymethynolide (10-dml), depending on whether PikAIV is bypassed (Figure 1.15). To confirm the quantities of each precursor produced, PikAIII and PikAIV were incubated with thiophenol pentaketide and methylmalonyl-CoA and 10-dml and nbl production were monitored. Production of 10-dml is a result of PikAIV having not modified the product, whereas nbl is produced if PikAIV has played its proposed role in the pathway. A number of chimeric PikAIII and PikAIV constructs were then produced with the C-terminal docking domain of PikAIII replaced by the ACPdd's from CurG and CurK and the N-terminal docking domain of PikAIV replaced by the KSdd's from CurH and CurL (Figure 1.16).

The two module throughput reaction was repeated with each combination of chimeric construct, resulting in both nbl and 10-dml production for all chimeras containing compatible docking domains, with around 50 % of nbl production retained. No nbl was detected for the majority of chimeras containing incompatible curacin docking domains. The one exception was between PikAIII with the KSdd of CurK (PikAIII-ddCurK) and PikAIV with the ACPdd of CurH (ddCurH-PikAIV), where around 25 % of nbl production was retained.

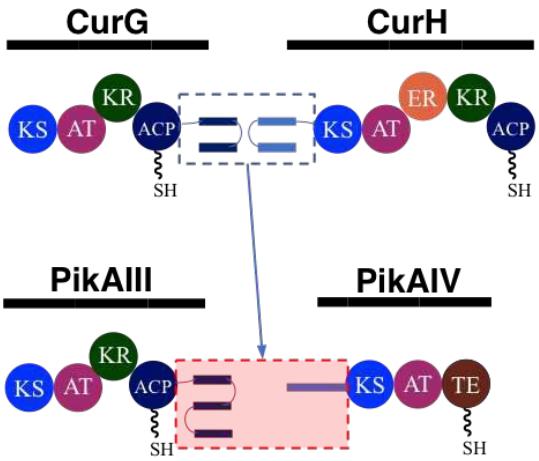


Figure 1.16. *The curacin PKS pathway (top) contains compatible class 2 docking domains to facilitate intersubunit interactions. Previous work has shown that these docking domains may be fused onto interacting modules from pikromycin PKS, in place of the natural class 1 pikromycin docking domains, to facilitate acyl transfer (bottom).*

The same reaction resulted in an increased level of nbl production (45 %) upon the mutation of CurK (E2224R). The mutation is likely to have created favourable electrostatic interactions between the docking domains, potentially with a glutamine in CurH's docking domain, demonstrating the importance of electrostatic interactions in docking domain selectivity. This highlights the possibility of changing electrostatic interactions through single point mutations, widening the possibility of engineering docking domains to ensure the compatibility of PKS subunits.

1.3 Spirotetronates

A subsection of polyketide metabolites are the spirotetronates. Tetronates contain a 4-hydroxy butenolide motif and are found in a wide variety of natural products, including ascorbic acid and terpenes.[135;136]

A macrocyclic compound containing a tetronic acid linked to a cyclohexene at a spiroatom is known as a spirotetronate.[137] Since their discovery in 1934,[138] over 100 naturally occurring spirotetronates have been identified.[28] Abyssomicins (Figure 1.17), a group of small spirotetronates (central ring system C_{11}), were first discovered in *Verrucosispora maris* AB 18-032, before the isolation of homologs in land-based soil.[139] Abyssomicin C is of particular interest due to observed bioactivity against *Mycobacterium tuberculosis*, Methicillin-

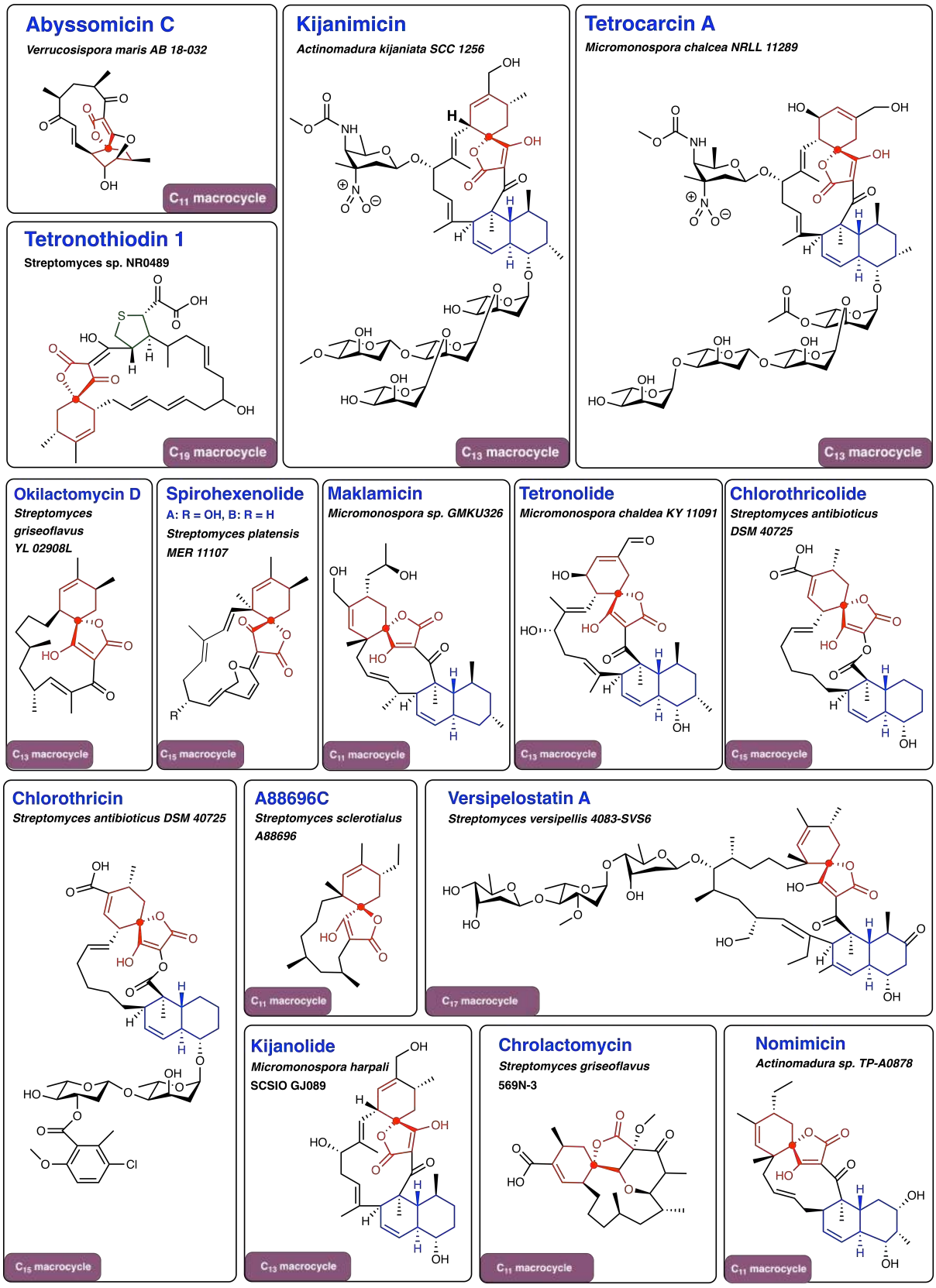


Figure 1.17. *Examples of Actinobacter spirotetronates with their spirocompounds and their spiroatoms shown. Medium spirotetronates (C_{13}) and large spirotetronates ($> C_{13}$) often contain either a trans-decalin or a tetrahydrothiophene moiety.*

resistant *Staphylococcus aureus* (MRSA) and the para-aminobenzoate pathway in *Bacillus subtilis*.[140]

Examples of medium spirotetronates (C_{13}) include kijanimicin, which is produced by *Actinomadura kijaniata*.[141] Kijanimicin inhibits the growth of Gram-positive bacteria, showing particularly strong activity against *Bacillus subtilis* and *Propionibacterium acnes*. Another medium spirotetronate that is closely related to kijanimicin is tetrocarcin A. Tetrocarcin A was isolated in 1980 from *Micromonospora chalcea* NRRL 11289 and exhibits anti-tumour activity, as well as interfering with anti-apoptotic pathways in lymphomas and human epithelial cells.[142] Tetronothiodin (C_{19}) is a large spirotetronate ($> C_{13}$), that was identified in *Streptomyces sp.* NR0489. Tetronothiodin contains a tetrahydrothiophene moiety in the position of the decalin-unit found in the medium spirotetronates and is a highly selective Cholecystokinin B (CCK_B) receptor antagonist, making it a valuable tool in the study of *CCK* receptors.[28]

1.3.1 The production of quartromicin precursors through a polyketide synthase pathway

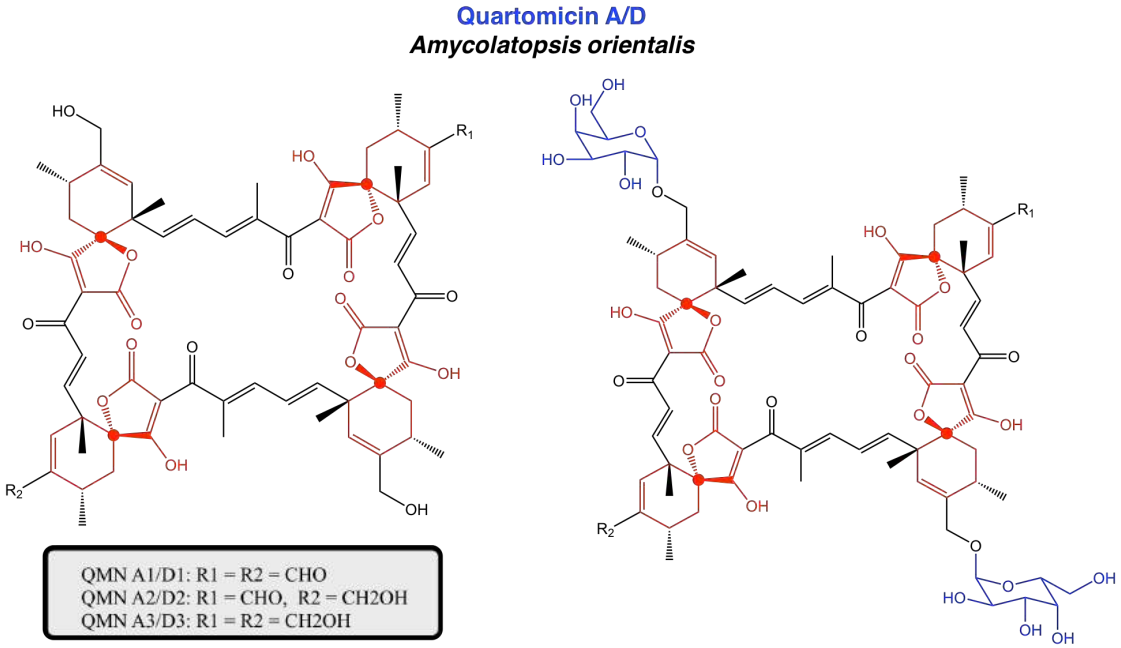


Figure 1.18. *The structure of quartromicin.*

Quartromicin (QMN), initially isolated from *Amycolatopsis orientalis* Q427-8, is a complex

of novel antiviral antibiotics containing the largest known spirotetronates (C_{32}).[143] Each antibiotic is unique due to the presence of four spirocompounds,[144] fused in a head-to-tail fashion by enone linkers, forming a carbon macrocycle. There are six known QMN components produced by the strain: A_1, A_2, A_3, D_1, D_2 and D_3 (Figure 1.18), which differ due to the presence of either an aldehyde or a hydroxymethyl group at R_1 and R_2 and the possible presence of galactose moieties. The glycosylated QMNs (QMN A_1 - A_3) showed significant antimicrobial activity against herpes simplex virus type I (HSV-1)[145] and the unglycosylated QMNs (QMN D_1 - D_3) showed significant activity against influenza.[146] In addition, QMN A_1 and QMN D_1 both demonstrated anti-HIV activity due to the inhibition of viral reverse transcription.[147;148]

As QMN has a wide range of biological activities, creating novel analogues could lead to the production of more effective pharmaceuticals. This provides the motivation to acquire a greater understanding of the control mechanisms involved in the production of QMN.

Previous unpublished work by members of the Challis group (Dr. Orestis Lazos and Dr. Malek Zerikly) had identified genes to encode most of the proposed proteins involved in the production of QMN. A less detailed gene cluster was subsequently published by another group.[147] The creation of genomic libraries for two quartromicin producing strains (*Amycolatopsis orientalis* ATCC 53884 and *Amycolatopsis mediterranei* NRRL 18815) and the amplification of a putative gene fragment from within the cluster allowed screening for the quartromicin gene cluster.[149] Shotgun sequencing and bioinformatic analysis of fosmids shown to contain the gene encoding an FkbH-like protein from the QMN biosynthetic pathway and overlapping genes led to the identification of genes to encode PKS modules, proteins to catalyse tetronate biosynthesis, oxidoreductase tailoring enzymes and a number of regulatory and resistance genes (Figure 1.19).

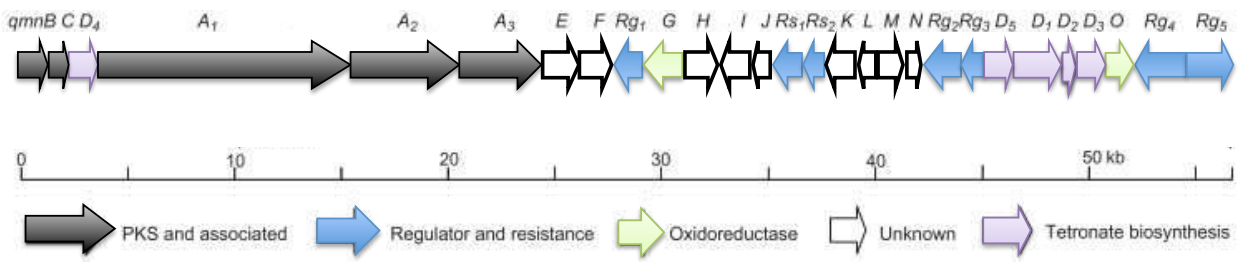


Figure 1.19. *The proposed gene cluster for QMN biosynthesis[147], including PKS and tetronate biosynthesis (Table 1.1).*

Table 1.1. *The QMN gene cluster and the proposed functions of encoded proteins.*

Gene name	Length (bp)	Proposed function of encoded protein
***qmn*B**	1,413	propionyl-CoA carboxylase
***qmn*C**	753	TE II
***qmn*D$_4$**	966	2-oxoacid dehydrogenase/AT
***qmn*A$_1$**	17,772	PKS (KS^Q-AT-ACP-KS-AT-DH-KR-ACP-KS-AT-DH-KR-ACP-KS-AT-DH-KR-ACP)
***qmn*A$_2$**	5,313	PKS (KS-AT-DH-KR-ACP)
***qmn*A$_3$**	3,876	PKS (KS-AT-KR^D-ACP)
***qmn*E**	1,236	unknown
***qmn*F**	1,488	unknown
***qmn*Rg$_1$**	255	regulator
***qmn*G**	1,599	PQQ-dependent dehydrogenase
***qmn*H**	1,128	Proposed Diels Alderase
***qmn*I**	1,044	unknown
***qmn*J**	483	unknown
***qmn*Rs$_1$**	1,176	transporter
***qmn*Rs$_2$**	681	transporter
***qmn*K**	1,056	unknown
***qmn*L**	456	unknown
***qmn*M**	657	unknown
***qmn*N**	390	unknown
***qmn*Rg$_2$**	1,347	regulator
***qmn*Rg$_1$**	666	regulator
***qmn*D$_5$**	1,029	3-oxoacyl-ACP synthase III (KS)
***qmn*D$_1$**	1,839	glyceryltransferase/phosphatase
***qmn*D$_2$**	213	ACP
***qmn*D$_3$**	843	2-oxoacid dehydrogenase/AT
***qmn*O**	1,209	cytochrome P450
***qmn*Rg$_4$**	2,805	regulator
***qmn*Rg$_5$**	2,463	regulator

*qmn*A$_1$-A$_3$ are proposed to encode three PKS subunits containing a loading and five extending modules, each containing an ACP, AT and KS domain, with four of the modules also containing KR and DH domains.[150] The loading module contains a KS^Q that likely functions as a malonyl-thioester decarboxylase to procure an acetyl starter unit.[147] The loading module, as well as modules 3 and 5, contains an extender unit binding motif in the AT domain with specificity towards a malonyl-CoA extender unit (HAFH),[72;151]

whilst modules 1, 2 and 4 contain extender unit binding motifs specific to methylmalonyl-CoA (YASH). Malonyl- and methylmalonyl-CoA are produced from acetyl- and propionyl-CoA by a propionyl-CoA carboxylase (*qmn*B).

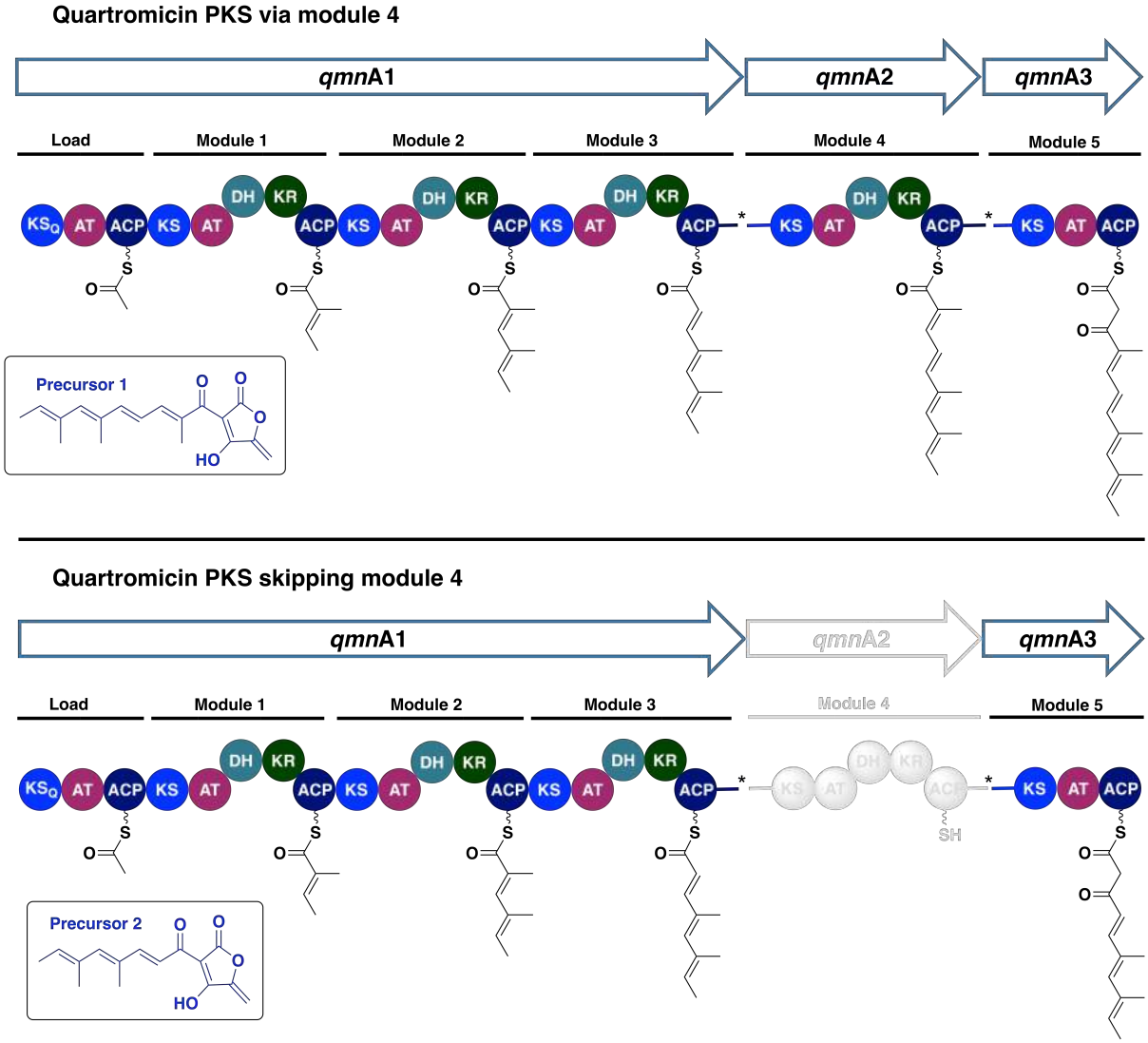


Figure 1.20. *The proposed biosynthetic pathways for quartromicin PKS[147] and the resulting precursors (**1** and **2**). Docking domains (*) facilitate intersubunit interactions between modules 3, 4 and 5.*

The production of QMN is proposed to require two different precursors.[147] Precursor **1** is the result of each PKS module sequentially extending and modifying the growing natural product. Precursor **2** is one extension shorter, probably as a result of module 3 being able to interact directly with module 5 and optionally bypass module 4 (Figure 1.20). The C-termini of module 3 and module 4's (M3 and M4) ACPs and the N-termini of the KSs (M4 and M5) appear to contain class I docking domains.

1.3.2 Spirotetronate formation

The mechanism for the formation of quartromicin can be predicted by comparison to the known roles of homologous enzymes in abyssomicin biosynthesis.[140;152;153] $QmnD_1$ is proposed to load the 3-carbon glycerol group of 1,3-biphosphoglycerate onto an ACP ($QmnD_2$), yielding glyceryl-S-ACP (Figure 1.21).[147;154] A stand-alone KS domain ($QmnD_5$) then catalyses bond formation between the glyceryl-S-ACP and the ACP domain-tethered polyketide chain of the final module. *qmn*D_3 is thought to encode an AT domain that acetylates the hydroxyl group of the tetronate ring, before the acetic acid is eliminated by a 2-oxoacid dehydrogenase/AT domain ($QmnD_4$) to form an exocyclic double bond.[155]

Abyssomicin only contains one spiroatom, as opposed to QMN's four. Therefore, genes involved in the subsequent steps of QMN biosynthesis could not be predicted from the abyssomicin gene cluster. Due to the structure of QMN's skeleton, is proposed that the cyclohexene moieties attached to tetronic moieties are formed by a two-step reaction involving two of each precursor. Firstly, an enzyme-catalysed [4+2] endo Diels-Alder reaction is predicted to occur between a diene (in the tail region of precursor **1**) and a dienophile (in the tetronic ring system of precursor **2**) to produce a homodimer. A [4+2] exo Diels-Alder reaction is then predicted to occur between the diene (in the tail region of precursor **2**) and dienophile (in the tetronic ring system of precursor **1**) of two homodimers to produce the QMN skeleton.[156] It is currently not known which protein(s) catalyses this reaction.

1.3.3 Modifications to the quartromicin skeleton

Following biosynthesis of the core QMN ring system, oxidation-glycosylation reactions further diversify the structure of QMN. A cytochrome P450 (QmnO) hydroxylates the methyl groups of the exo Diels-Alder adducts (Figure 1.22). QmnG, a pyrroloquinoline quinone (PQQ) dependent oxidoreductase, then catalyses the dehydrogenation of the hydroxymethyl groups into aldehydes. The galactose moieties that differentiate quartromicin A from quartromicin D may be transferred from a nucleotide diphosphate to the hydroxy-methyl groups of the endo Diels-Alder adducts by a glycosyltransferase encoded by an unidentified gene.[157]

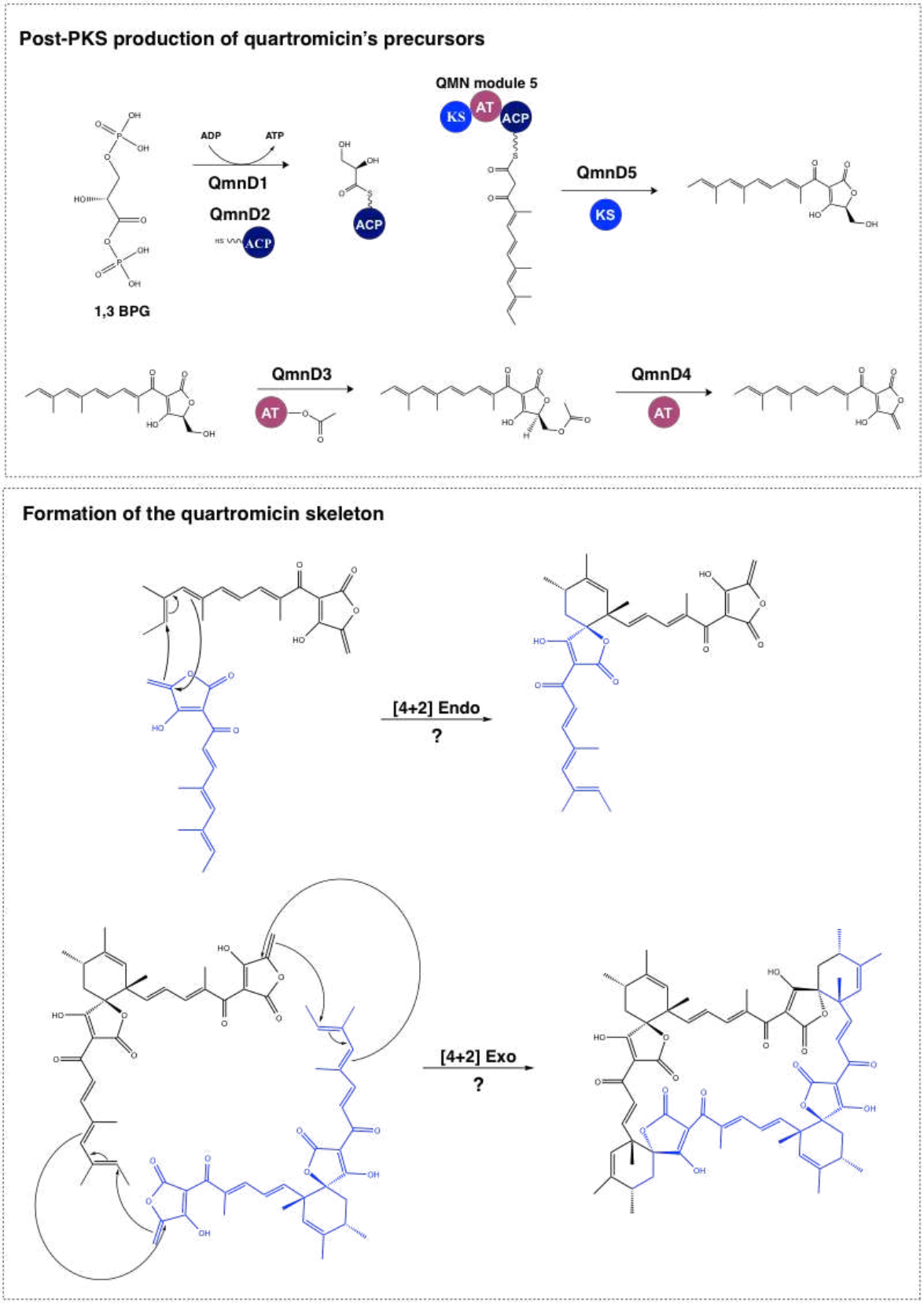


Figure 1.21. *The proposed formation of QMN from the PKS precursors.*

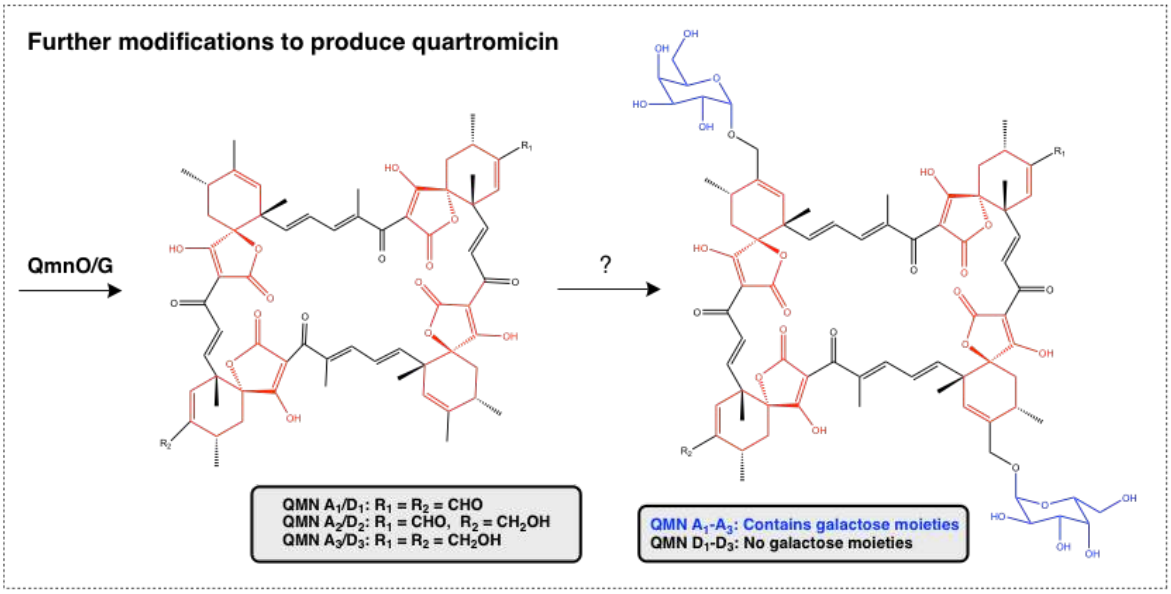


Figure 1.22. *The proposed modification of QMN's skeleton to form the six bioactive components of QMN.*

1.3.4 Regulatory and unknown genes

Although *qmn*E shows close identity to a gene encoding a β-lactamase and *qmn*F to a gene encoding a pyridoxal phosphate-dependent cysteine desulfurase, both have unknown roles.[155] Five genes (*qmnRg*$_1$-*Rg*$_5$) encode putative regulatory proteins. *qmnRg*$_1$ encodes a ChlF2-like regulator, *qmnRg*$_2$ and *qmnRg*$_3$ encode two-component regulators, *qmnRg*$_4$ encodes a LuxR family regulator and *qmnRg*$_5$ encodes a *Streptomyces* antibiotic regulatory proteins (SARP) family activator. *qmn*C encodes a type II TE that is predicted to be responsible for removing incorrect PKS intermediates. In addition, *qmnRs*$_1$ and *qmnRs*$_2$ encode ATP-binding casette (ABC) transporters and *qmn*I, *qmn*J, *qmn*K, *qmn*L, *qmn*M and *qmn*N encode proteins with unknown functions.

1.4 Aims and objectives

The overall aim of this work was to understand the molecular interactions that control the proposed module skipping in QMN biosynthesis, which allows the KS domain of module 5 to accept substrates both from the ACP of module 3 and module 4 and catalyse a decarboxylative Claisen condensation between either substrate and a malonyl-CoA extender unit. This is important because experiments that try to incorporate non-natural PKS domains have often been unsuccessful, likely due to a lack of understanding of how

protein-protein interactions work within PKSs.

The first aim was therefore to overproduce and purify the ACP domains from modules 3 and 4, with and without their docking domains, as well as module 5, both as a full-length module and with various truncations. The objective was then to probe the intramodular interactions within QMN module 5. To do this, it would first be necessary to demonstrate the catalytic activity of the complete module by intact mass spectrometry.[158;159] Subsequent similar experiments with a variety of substrate analogues were planned to probe the selectivity of the KS and AT domains[160] before intramodular chemical cross-linking experiments between the intramodular ACP and the KS/AT domains were planned to identify which domains in module 5 would interact together.[161–163]

The second aim was to attempt to probe the molecular interactions within module 5 through structural elucidation via techniques such as gel filtration, NMR, X-ray crystallography and cryo-EM.[164;165]

The third and final aim was to investigate the intermodular interactions between module 5 and the upstream ACP domains of modules 3 and 4, as well as the role of the docking domains in this interaction. This was intended to be studied by intact mass spectrometry assays to monitor the rate of transfer between the upstream ACP domains, both with and without the docking domains, and module 5. The importance of each domain within module 5 in facilitating intermodular interactions and the strength of the binding was intended to be investigated by cross-linking and SPR, respectively. In addition, structural investigation of the ACP docking domains of module 3 and 4 and the KS docking domain of module 5 was proposed by NMR and X-ray crystallography.

Chapter 2

Overproduction, purification and characterisation of the recombinant proteins involved in quartromicin's unique module skipping mechanism

2.1 Target proteins

Understanding the mechanisms involved in the production of natural products would provide the possibility of systematically developing synthetic pathways to produce novel natural products, as well as modifying existing systems or creating hybrids. This would allow the production of bespoke analogues designed with specific functionalities in mind. One way to gain an insight into the enzymatic workings of a PKS pathway is by examining the intermodular interactions between the subunits, as well as the intramodular workings of modules. *In vitro* studies are often used to investigate PKSs due to issues culturing organisms under *in vivo* conditions. Prerequisites for studying protein-protein interactions using *in vitro* bioassays are to overproduce and purify the relevant recombinant proteins and produce mimics of the proteins' substrates.

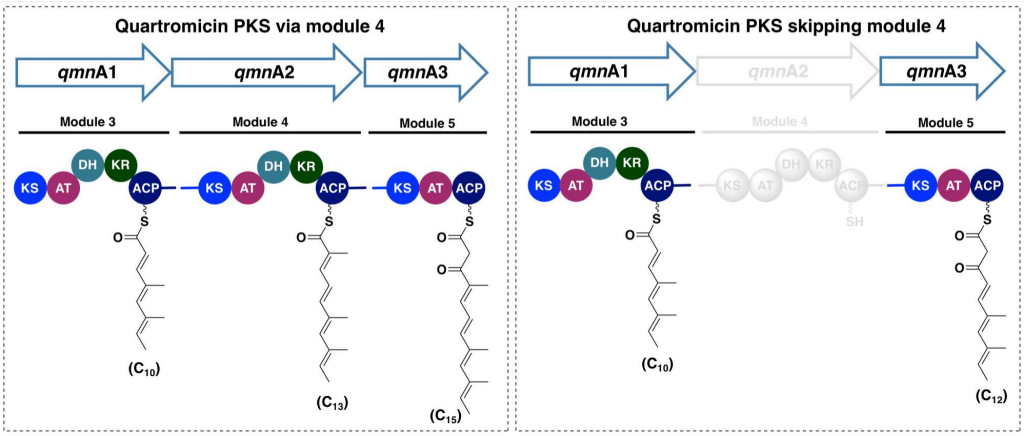


Figure 2.1. *QMN module 5 is able to interact with the upstream ACPs from modules 3 and 4, allowing the bypassing of module 4.*

To elucidate information on the module skipping mechanism of QMN PKS, the interactions between the domains and modules involved in the process were studied. The intermodular interactions between the ACP domains of modules 3 (M3) and 4 (M4) and the KS domain of module 5 (M5), as well as the intramodular interactions within M5, may play key roles in moderating the specificity of the PKS pathway (Figure 2.1). The ACP domains from M3 and M4 both contain a four conserved helices containing the ACPs substrate attachment site, as well as a C-terminal docking domain connected to the four conserved helices by a largely unstructured linker region. The ACP docking domains are reported to facilitate interactions between the ACP and the KS of the subsequent module, which contains an N-terminal docking domain hypothesised to have a specificity to interacting proteins.[124] Two different variations of each ACP were targeted, a full-length

protein (M3_ACP and M4_ACP) and a truncated variant lacking the docking domain and linker region (M3_ACPT and M4_ACPT). This would allow the elucidation of information on the importance of each docking domain in facilitating protein-protein interactions between the ACP and the downstream KS domain and determine whether the ACP's four conserved helices or the docking domain plays the primary role in the specificity of ACP-KS interactions. Homology models were used to determine the start of the disordered linker region (Figure 2.2).

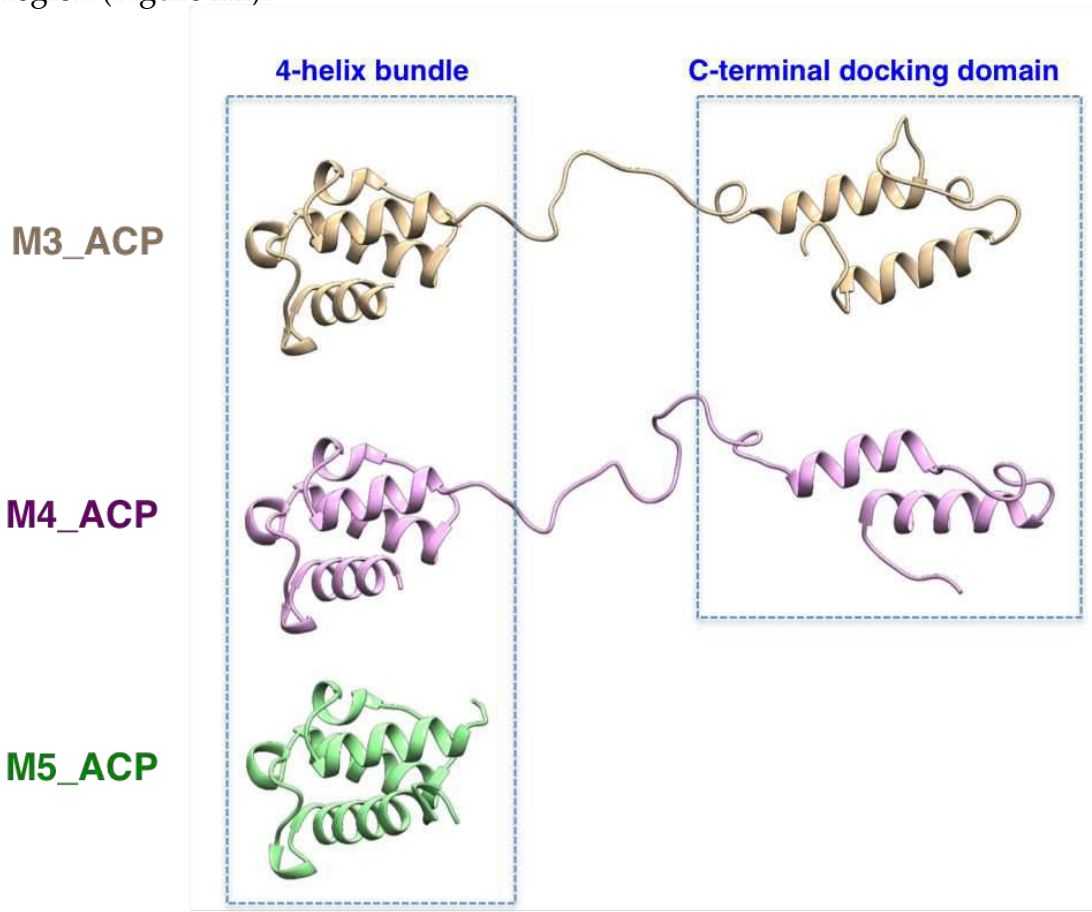


Figure 2.2. *Homology models (Phyre, PDB: 2ju2, 5hvc) of the QMN ACP domains from the final 3 modules of QMN PKS. M3_ACPT and M4_ACPT consist of the residues in the 4 conserved helices.*

As well as the full-length M5, a number of truncated constructs were planned. The stand-alone KS domain (M5_KS), KSAT didomain, both with (M5_KSAT) and without (M5_KSATS) the AT-ACP linker, and the ACP (M5_ACP) were all targeted to allow for determination of the effects of each region on protein-protein interactions (Figure 2.3).

M5_KS contains the docking domain with specificity to the upstream ACPs and an archetypal KS domain containing the active-site cysteine to which the upstream modules polyketide substrates are transferred. Therefore, excising the individual domain will allow for assays to demonstrate that the protein-protein interactions are specific to the KS and ACP.

	Contains	Residues
M5_KS	KS domain	1-464
M5_KSATS	KS-AT didomain	1-888
M5_KSAT	KS-AT didomain (with linker)	1-1198
M5	Intact module	1-1292
M5_ACP	ACP domain	1205-1292

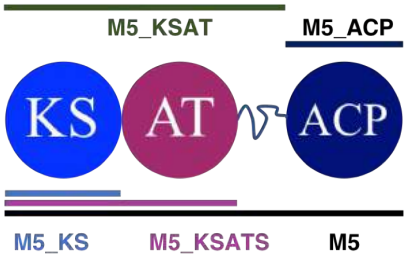


Figure 2.3. *M5_KS (1,392 bp), M5_KSATS (2,664 bp), M5 _KSAT (3,594 bp) and M5_ACP (264 bp) were targetted for subcloning from M5 (3,876 bp).*

Subcloning of M5_KSATS and M5_KSAT will allow for studies into the importance of both the AT and AT-ACP linker region (which is predicted to contain structured KR remnants, including a partial Rossmann fold) in interactions between domains, in particular with M5_ACP and the upstream ACPs. M5_ACP is at the C-terminal of QMN PKS, and therefore only consists of a four conserved helices due to the lack of necessity in interacting with a downstream KS. Therefore, only one variant of M5_ACP will be targeted.

As well as being useful for determining which regions are important for protein-protein interactions, the creation of different lengths of truncated constructs will aid structural studies. Whilst full-length M5 could be studied by cryogenic-electron microscopy, the resolution of these structures is generally lower than structures obtained from X-ray crystallography or NMR. The creation of the aforementioned truncated variants, specifically variants lacking flexible regions, will aid analysis using structural biology techniques.

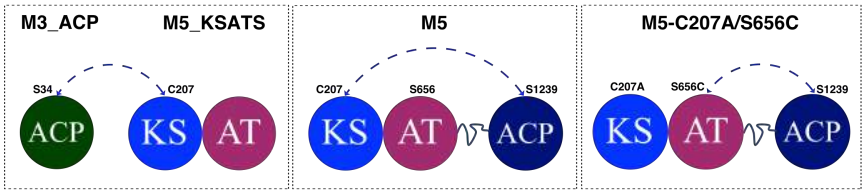


Figure 2.4. *Cross-linkable interactions occur between M5_KS and the upsteam ACPs (M3_ACP and M4_ACP), as well as the intramodular ACP (M5_ACP). Cross-linkable interactions should also be present between M5's AT and M5_ACP upon converting the AT's active-site to a cysteine (S656C).*

In addition to the truncated variants of M5, constructs containing active-site mutations were targeted. Mutation of the KS domain's active-site cysteine to an alanine (C207A) has previously been shown to abolish the activity of KS domains,[166] thus acting as a vital control for *in vitro* bioassays involving M5. A second mutation of the AT's active-site serine to alanine (C207A/S656C) was planned in order to facilitate interactions between the AT and ACP domains and allow *in vitro* studies to be undertaken (Figure 2.4).

2.2 Plasmid preparation

Plasmids for the overproduction of M3_ACP, M4_ACP, M5_KSAT and M5 were previously produced by Dr. Orestis Lazos and Dr. Matthew Jenner. All other plasmids were prepared by PCR amplification of the GOI and ligation into the vector backbone, or through modification of an existing plasmid.

2.2.1 Plasmid preparation through PCR amplification and ligation

Each targeted gene of interest (GOI) was amplified by polymerase chain reaction (PCR), with forward and reverse primers designed to be complementary to 15 - 18 bases at the 5′ and 3′ ends of the GOI. The forward primer also contained additional bases (CACC) at the 5′ end to facilitate ligation to the complementary sticky end of the host vector (GTGG), ensuring the correct orientation of the inserted GOI. Where a required stop codon was not already present in the GOI, an artificial in-frame stop codon was added to the 5′ end of the reverse primer (Figure 2.5).

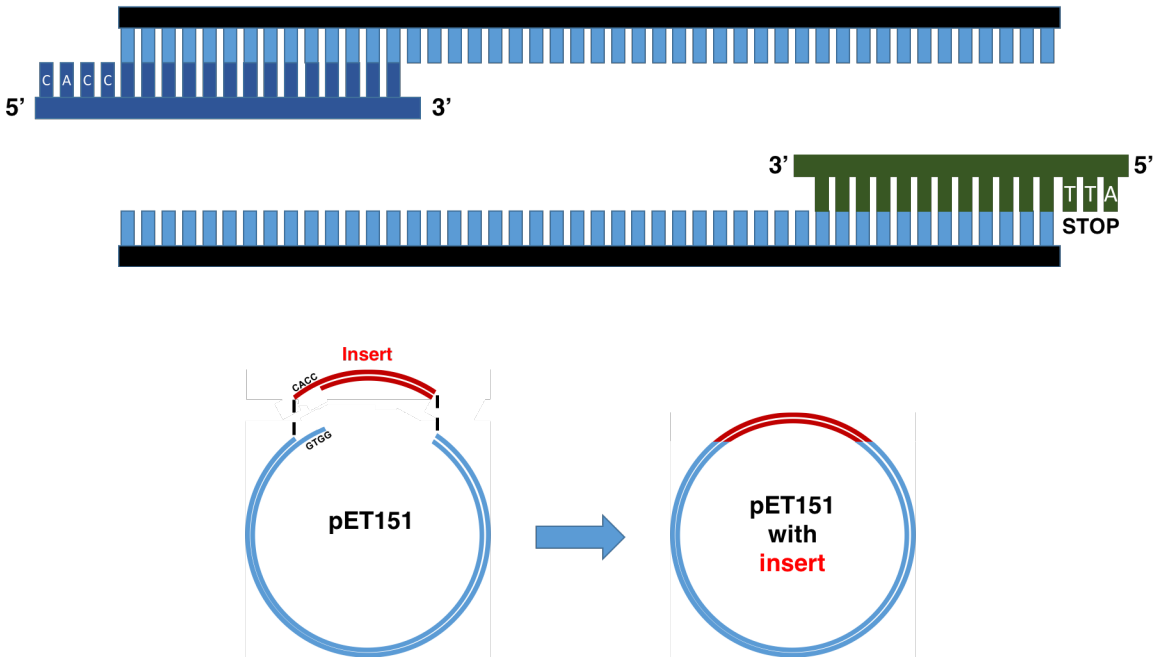


Figure 2.5. *Primers encode a CACC overhang on the 5′ end of the replicated region (top) to ensure the PCR product has the correct orientation after ligation into pET151 (bottom).*

PCR products were visualised by agarose gel electrophoresis and bands corresponding to the expected sizes were excised and purified (Figure 2.6). Sufficient quantity of PCR product was produced for the genes corresponding to M3_ACPT, M4_ACPT, M5_ACP and M5_KS, but reactions were unsuccessful for M5_KSATS and M5_KSAT, perhaps due to the length of the GOI.

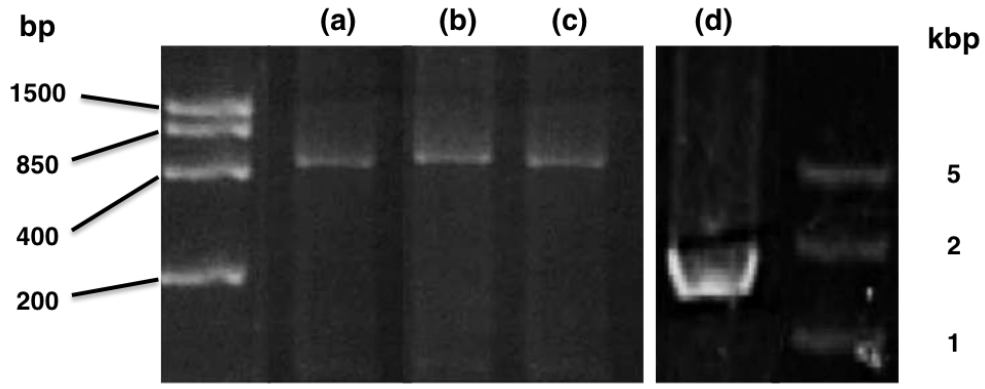


Figure 2.6. *Agarose gel electrophoresis following PCR amplification of genes encoding M3_ACPT (**a**, 247 bp) M4_ACPT (**b**, 238 bp), M5_ACP (**c**, 268 bp) and M5_KS (**d**, 1,405 bp).*

The insertion of the purified GOIs into a pET151 vector was achieved using an expression kit containing a linearised topoisomerase I-activated pET151-D/TOPO vector. Topoisomerase I is an enzyme from *Vaccinia* virus that cuts a single strand of double stranded-DNA, facilitating strand relaxation. The 5′ end of the GOI is then able to anneal to the complementary sticky end of the vector whilst the 3′ end of the GOI anneals to the 5′ end of the vector. This avoids the necessity of multiple DNA ligation reactions and therefore can be completed in a relatively short time frame. The ligated product was then used to transform *E. coli* and plasmid-containing cells were selected through LB agar ampicillin plates. Colonies were picked and grown overnight in LB.

Grown cultures were centrifuged to remove the media from the cells, before plasmid DNA was extracted and purified. Restriction enzymes with one restriction site in both the vector and the insert were identified and digests were carried out on the purified DNA plasmids. Restriction digested clones displaying the correct band sizes following agarose gel electrophoresis (Figure 2.8) were sent for sequencing to determine their fidelity. This confirmed successful cloning of the genes encoding M3_ACPT, M4_ACPT, M5_ACP and M5_KS, respectively.

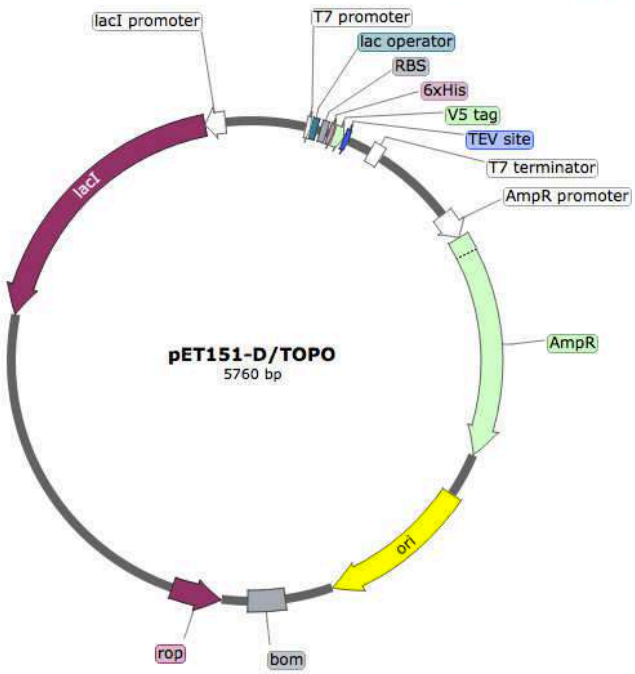


Figure 2.7. *A gene map of the pET151 D/TOPO vector (SnapGene).*

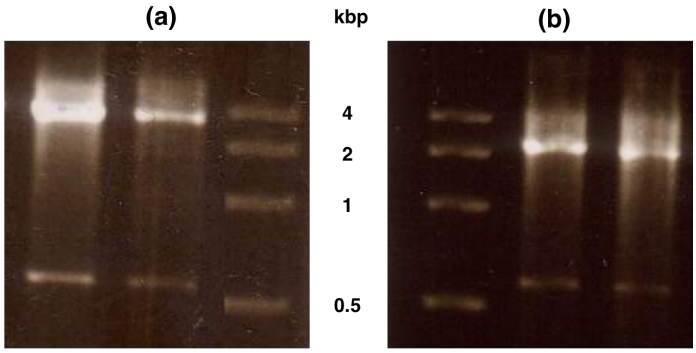


Figure 2.8. *Agarose gel electrophoresis following restriction digests of M3_ACPT (**a**, expected band size: 1,015 and 4,985 bp) and M5_ACP (**b**, expected band size: 1,120 and 4,880 bp) by PvuI.*

2.2.2 Plasmid preparation through site-directed deletion

Due to problems ligating large PCR fragments into the pET151 vector, a new approach was chosen to subclone the genes encoding M5_KSATS and M5_KSAT by deleting the relevant regions of the plasmid. Non-mutagenic primers were designed, using NEBasechanger, to initiate copying from the flanking regions on either side of the required deletion (Figure 2.9 (left)).

Primers were designed back-to-back to ensure exponential amplification of the product, as well as to transform non-nicked plasmids. A series of stop codons were also included at the 3′ end of the reverse primer to prevent the transcription of extra bases. Clones were

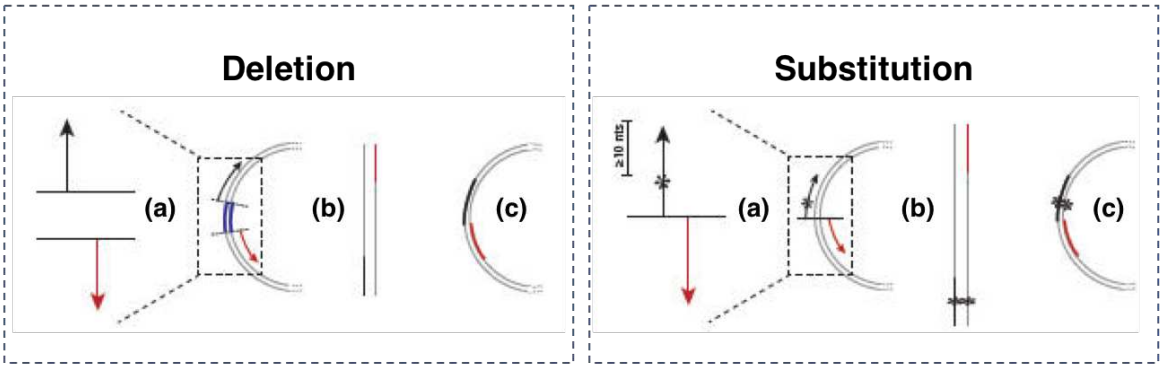


Figure 2.9. *Site-directed mutagenesis and deletion can be carried out on a plasmid. Primers complementary to the GOI, but with either the omission of (left), or changes (right), a number of bases **(a)**, allow for the production of a PCR product with the required deletion or base changes, respectively **(b)**. In both cases, the PCR product may be ligated, producing the modified plasmid **(c)**.*

checked by restriction digest and sequenced, confirming the successful deletion of the relevant regions of the genes encoding M5_KSATS and M5_KSAT.

2.2.3 Site-directed mutagenesis of quartromicin module 5 constructs

All site-directed mutagenesis (SDM) was carried out using a SDM kit that provides the necessary elements for the successful production of PCR products containing the required base pair mutation(s) and subsequent ligation of the plasmid (Figure 2.9 (right)). Primers were designed back-to-back using the NEBasechanger and OligoAnalyzer was used to confirm no hairpins or dimerisation would occur between the primers. The KS active-site of the genes encoding M5 and M5 truncations would first be mutated from cysteine to alanine. Then a mutation on the active-site of the AT domain would be carried out to produce double mutants of each gene.

The PCR reactions for the C207A mutations were analysed by agarose gel electrophoresis and reactions exhibiting bands of the required length were treated with the provided KLD (kinase, ligase, DpnI) enzyme mix to ensure phosphorylation of the 5' ends of the PCR product, circularisation of the plasmid and digestion of methylated template DNA. The treated plasmids were then transformed, cultured and isolated from the bacteria. Sequencing confirmed no mutation had occurred for M5 or M5_KSAT. However, the mutation was successful in M5_KSATS (M5_KSATS-C207A). A second mutation (S656C) was also successful for M5_KSATS-C207A, yielding the double mutant (M5_KSATS-C207A/S656C). After numerous unsuccessful attempts, it was hypothesised that M5 and M5_KSAT

in the pET151 vector were too long to introduce the mutation by this method.

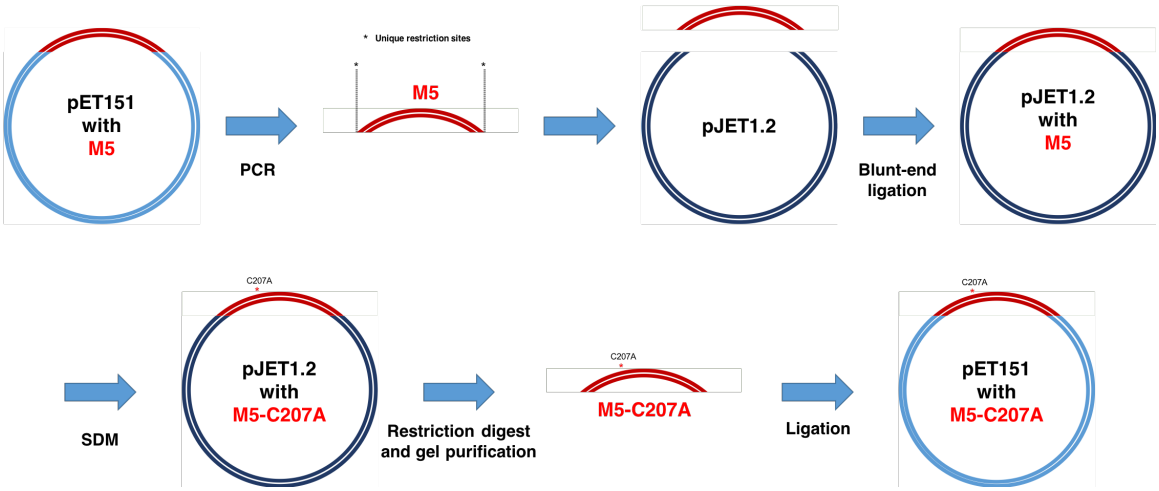


Figure 2.10. *PCR products containing restriction sites at both ends may be blunt-end ligated into a pJET vector, allowing for more efficient mutagenesis of the insert. The mutated insert may then be isolated from the vector by restriction digest and gel purification, before ligation back into pET151.*

Two parallel methods were attempted to overcome the inefficacy of the mutations in M5_KSAT and M5. The first method involved transferring M5 into pJET1.2, a shorter cloning vector (2974 bp), and carrying out the mutations on the shorter construct (Figure 2.10). The primers to copy the GOI contained restriction sites at the 5′ and 3′ ends to allow removal of the mutated GOI from pJET1.2 and ligation back into pET151.

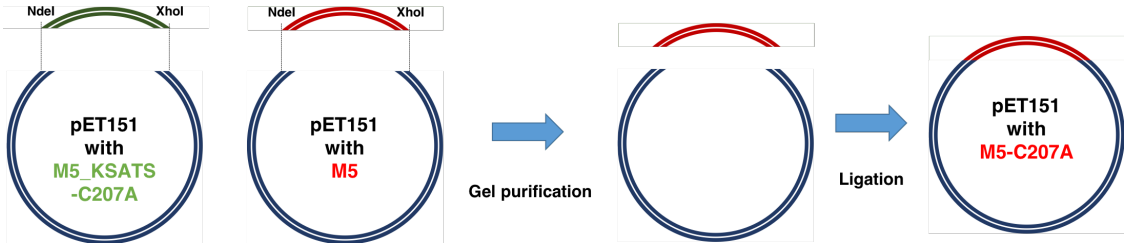


Figure 2.11. *The region of M5_KSATS containing the mutations may be restriction digested out of its vector and purified alongside the complementary region of M5. Ligation of the two regions may be achieved by incubation with T4 ligase, resulting in the production of a plasmid encoding M5-C207A.*

The second method involved identifying unique restriction-sites in M5_KSATS (pET151) that would allow the separation of a region containing the mutations from the rest of the construct. NdeI was chosen as the only predicted cleavage was in the multiple cloning site (MCS). XhoI was also identified as cutting the plasmid once, with its restriction-site located after the second mutation (S656C), but before the end of the AT domain. This would allow the excision of the area containing the mutated KS and AT active-sites from M5_KSATS-C207A and M5_KSATS-C207A/S656C. It would also allow the complementary region of M5 to be excised. This contained the remainder of the AT domain, the

AT-ACP linker, M5_ACP and the majority of the pET151 vector. Ligation of the regions containing the mutated KS and AT active sites to the complementary region of M5 would allow for the creation of M5-C207A and M5-C207A/S656C (Figure 2.11).

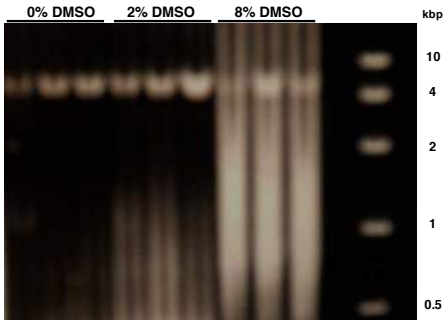


Figure 2.12. *An agaorse gel of PCR products for M5-C207A in the pJET1.2 vector (expected band size: 6,706 bp).*

PCR of M5 containing 5′ and 3′ restriction sites was successful and the product was inserted into pJET1.2 by blunt end ligation (confirmed by sequencing). PCR with the mutagenic primers yielded a product that was inserted into the vector. Sequencing confirmed the successful C207A mutation (Figure 2.12).

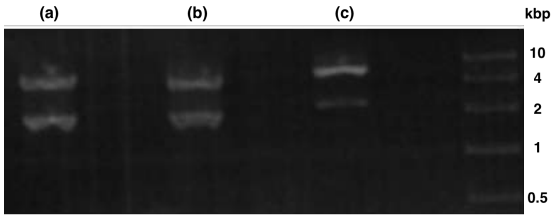


Figure 2.13. *Agaorse gel electrophoresis of a restriction digest of M5* ***(a)****, M5_KSATS-C207A* ***(b)*** *and M5_KSATS-C207A/S656C* ***(c)*** *with NdeI and XhoI. Expected band sizes are 2.0/6.4 kbp for M5_KSATS-C207A and M5_KSATS-C207A/S656C and 2.0/7.4 kbp for M5.*

At the same time, agarose gel electrophoresis suggested successful double restriction of M5_KSATS-C207A, M5_KSATS-C207A/S656C and M5 had occurred (Figure 2.13). The required polynucleotides were excised and purified by gel extraction. Ligation of the region containing the single mutant (C207A) and the double mutant (C207A/S656C) to the polynucleotide containing most of the vector and the remainder of M5 yielded products that were confirmed to be of the correct size by restriction digest (Figure 2.14).

Fidelity of M5-C207A and M5-C207A/S656C were confirmed by sequencing. M5_KSAT variants of both plasmids and a C207A variant of M5_KS were then produced by the addition of in-frame stop codons and deletion of the relevant regions from M5-C207A and M5-C207A/S656C (Figure 2.15).

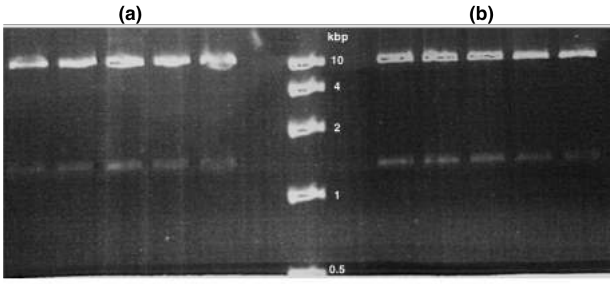


Figure 2.14. *An agaorse gel of a restriction digest of M5-C207A* ***(a)*** *and M5-C207A/S656C* ***(b)*** *with PstI. Expected band sizes were 1.2 and 8.4 kbp.*

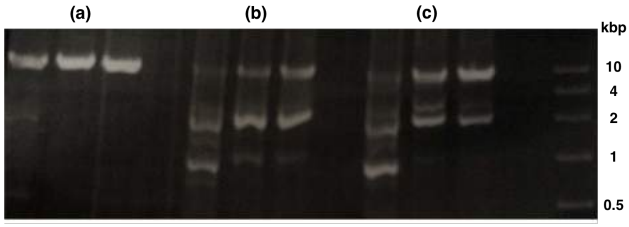


Figure 2.15. *An agaorse gel of PCR products for M5_KS-C207A (**a**, 7.2 kbp), M5_KSAT-C207A (**b**, 9.7 kbp) and M5_KSAT-C207A/S656C (**c**, 9.7 kbp).*

2.3 Overproduction and purification of proteins

2.3.1 Overproduction

Plasmids with the correct inserts were used to transform BL21* (DE3), a strain of *E. coli* that contains T7 RNA polymerase, allowing transcription of the plasmid's insert. Single colonies were selected on LB agar plates containing ampicillin and used to grow LB cultures until the cells reached logarithmic phase.

The addition of an allactose analogue, such as isopropyl β-D-1-thiogalactopyranoside (IPTG), allosterically released the repressor from the vector's T7 promoter site, allowing transcription of the GOI. An advantage of using IPTG over allactose is that it cannot be metabolised by *E. coli*, ensuring an invariable rate of expression. The *E. coli* ribosome was then able to bind to the RNA transcripts' ribosome binding site (RBS), allowing translation to be initiated by the GOI's start codon (ATG) and ended by a stop codon (TAG, TAA or TGA). This resulted in heterologous overproduction of the target proteins inside the *E. coli* host.[167]

Table 2.1. *The size (residues) and expected molecular weights (Da) of overproduced His-tagged proteins.*

Protein domain	M3_ACP (*)	M3_ACPT	M4_ACP	M4_ACPT	M5_ACP	M5_KS	M5_KSATS	M5_KSAT
Size (residues)	181	107	181	108	121	497	921	1231
Expected mass (Da)	19,608.83	12,212.79	19,465.05	12,019.77	12,915.41	52,862.21	97,887.79	131,211.28

M3_ACP, M3_ACPT, M4_ACP, M4_ACPT, M5_ACP, M5_KSATS, M5_KSAT and M5 were all successfully overproduced and were soluble in BL21* ^{TM} *E. coli* (Table 2.1). M5_KS overproduction initially resulted in a completely insoluble protein fraction, but a reduction in IPTG concentration (from 500 to 200 μM) resulted in the overproduction of soluble protein. All mutagenic variants displayed similar levels of overproduction to the corresponding wild type proteins.

2.3.2 Purification by nickel column

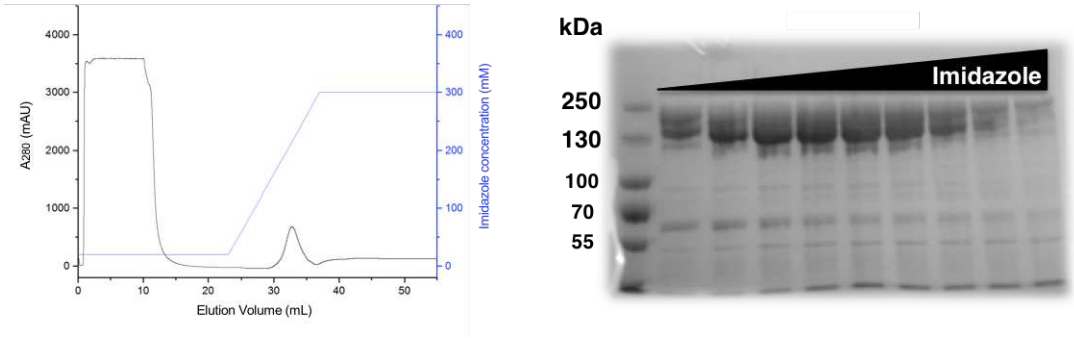


Figure 2.16. *IMAC chromatogram (left) and SDS-PAGE analysis (right) of M5 (141 kDa).*

Immobilised metal affinity chromatography (IMAC) was used as a first purification step for all soluble proteins (Figure 2.16), resulting in the successful purification of M3_ACP, M3_ACPT, M4_ACP, M4_ACPT, M5_ACP, M5_KS, M5_KSATS, M5_KSAT, M5 and all mutagenic variants (Figure 2.17). M5_KS also required the addition of a protease inhibitor (cOmplete^{TM}, EDTA-free Protease Inhibitor Tablets) to prevent degradation from occurring. M3_ACP co-expressed with a protein of higher molecular weight.

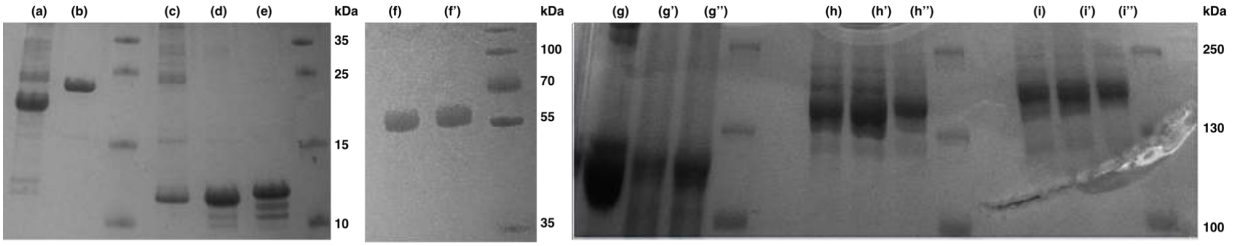


Figure 2.17. *SDS-PAGE analysis of His-tagged M3_ACP (**a**, 19,608 Da), M4_ACP (**b**, 19,465 Da), M3_ACPT (**c**, 12,213 Da), M4_ACPT (**d**, 12,020 Da), M5_ACP (**e**, 12,915 Da), M5_KS (**f**, 52.9 kDa), M5_KSATS (**g**, 97.9 kDa), M5_KSAT (**h**, 131 kDa) and M5 (**i**, 141 kDa) were produced and purified by IMAC. C207A (') and C207A/S656C mutants (") are also shown for constructs containing the KS and AT domains.*

2.3.3 Purification by gel filtration

Further purification steps were required for M5_KSATS, M5_KSAT and M5 in order to achieve a suitable level of purity for high-resolution structural studies. Therefore, size exclusion chromatography was used on the IMAC purified samples.

A Superdex 200 gel filtration column was calibrated using a number of analytical standards, producing a calibration curve from which the mass corresponding to each elution peak could be approximated (Figure 2.18). For all samples, some aggregated protein was eluted at the dead volume, perhaps as a result of the high protein concentrations prior to column injection. For M5_KSAT and M5, further peaks were observed, corresponding to masses of approximately 250 and 280 kDa, respectively, indicating both proteins formed dimers.

However, for M5_KSATS, a second peak was not visible, suggesting almost complete aggregation of the protein (Figure 2.19 (a)). The addition of protease inhibitor (cOmplete^{TM}, EDTA-free Protease Inhibitor Tablets) and increasing the salt concentration (200 mM from 100 mM) resulted in the presence of new peaks corresponding to the monomer, dimer and various oligomers (Figure 2.19 (b)). Further addition of 10 % glycerol resulted in the production of mostly monomeric protein (1.7 $\frac{V_e}{V_0}$), but large amounts of aggregation remained (Figure 2.19 (c)). The amount of aggregate was finally reduced by the addition of 100 mM EDTA (Figure 2.19 (d)). Calibration predicted a mass of approximately 100 kDa, suggesting M5_KSATS existed as a monomer. The loss of dimerisation in the truncated construct containing only the KS and AT domains (M5_KSATS) indicates that a dimerisation motif may be present in the AT-ACP linker region. This observation will be further studied in section 4.5.

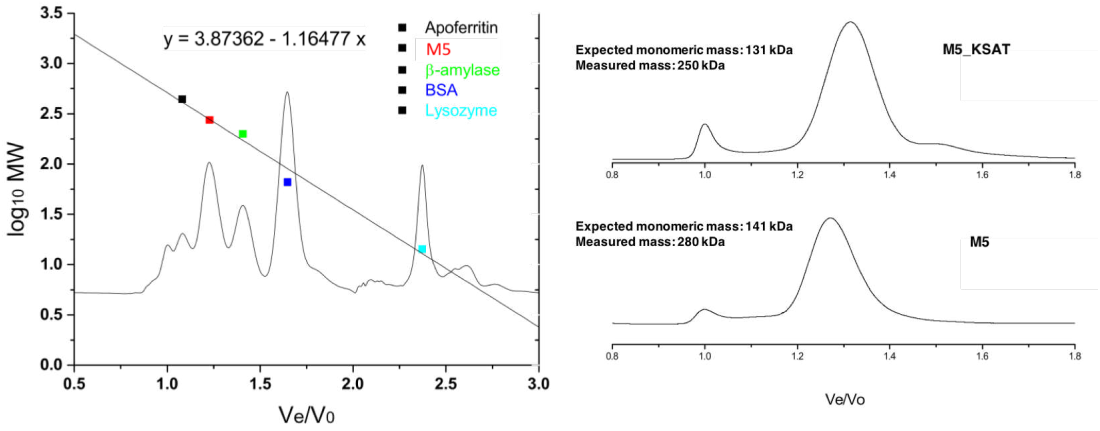


Figure 2.18. *A calibration curve was produced using proteins of known molecular weight (left, S200) and standardised by the elution time of thyroglobulin (V_0). Gel filtration analysis (S200) of M5_KSAT and M5 resulted in elution peaks of approximately 1.3 $\frac{V_e}{V_0}$ (top right) and 1.25 $\frac{V_e}{V_0}$ (bottom right), corresponding to molecular weights of approximately 250 kDa and 280 kDa, respectively.*

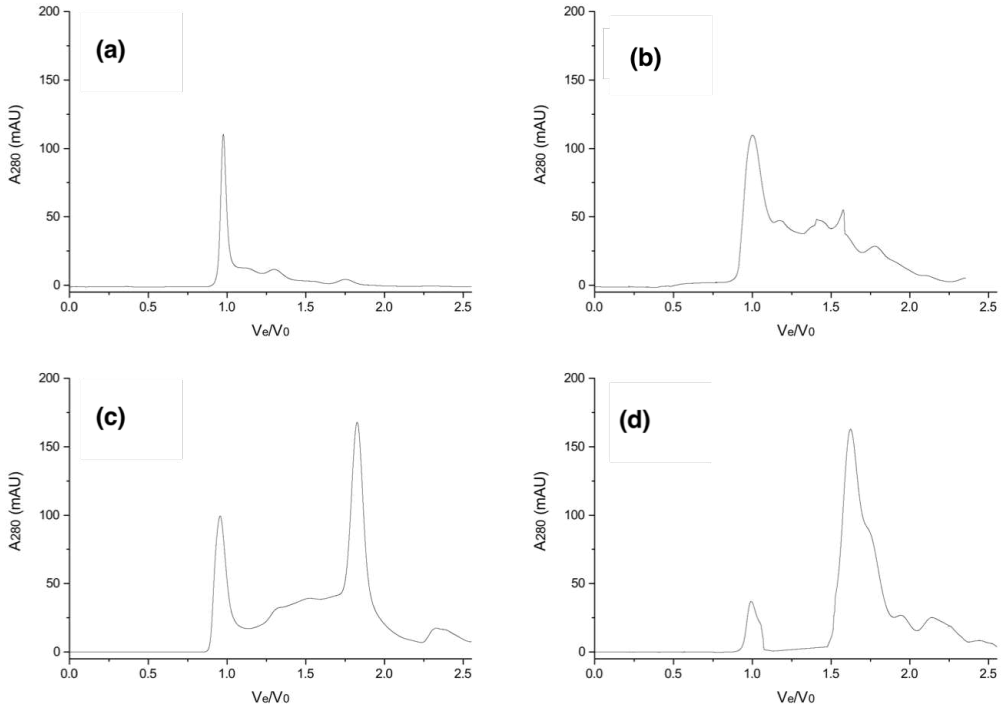


Figure 2.19. *Gel filtration analysis (S75) of M5_KSATS in different buffers:* ***(a)*** *20 mM Tris, 100 mM NaCl, pH 7.4,* ***(b)*** *+ 100 mM NaCl and protease inhibitors,* ***(c)*** *+ 10 % glycerol and* ***(d)*** *+ 100 mM EDTA.*

2.3.4 Estimating protein concentration

In order to ensure all assays were quantitatively accurate, the concentration of proteins must be accurately measured. To predict the concentration of each protein, a spectropho-

tometer was used to measure the absorbance of the sample at 280 nm. The protein concentration (C, mg/mL) was then estimated using the Beer-Lambert law:

$$C = \frac{A_{280}}{\epsilon l},$$

where A_{280} is the absorbance of the sample at 280 nm, ϵ is the protein's extinction coefficient and l is the path length of the cell. The extinction coefficient was calculated from the number of each aromatic residue (tryptophan, tyrosine, phenylalanine) and corresponds to how strongly the protein attenuates to light at 280 nm. Molarity was calculated by dividing the concentrations by the expected molecular weight of the protein.

All proteins were diluted to 0.5 mg/mL and a Bradford assay was used to confirm the accuracy of the measurement. A Bradford assay predicts protein concentrations based upon the absorbance shift of an anionic dye upon protein binding.[168] The absorbance shift can then be compared to the shift of a protein of known concentration, often bovine serum albumin (BSA). The anionic dye forms a non-covalent complex with the protein's carboxyl groups by donating its free electrons to the protein's ionisable groups. This causes disruptions to the proteins native state, exposing hydrophobic pockets that then bind non-covalently to the dye via Van der Waals forces. The proteins positively charged primary amines may then interact and bond to the negatively charged dye, causing the neutralisation of the dye and a change in colour that can be measured by the absorbance reading (595 nm). An advantage of this method is that it is not susceptible to various chemical compounds often found in protein samples, such as salts.

A calibration curve was measured with varying concentrations of BSA (Figure 2.20), before the absorbance was measured for each protein. Comparisons with the calibration curve predicted concentrations of 0.476 and 0.501 mg/mL for M3_ACP and M5 (as expected), and concentrations of 0.188, 0.284 and 0.359 mg/mL for M3_ACPT, M4_ACP and M4_ACPT, respectively. This variance in predicted concentration could be due to a lack of residues with absorbance at 280 nm, resulting in an inaccurate extinction coefficient leading to an overestimation of concentration by nanodrop. As a result of this disparity in ACP protein concentrations, it was decided to lyophilise samples of each ACP, thereby allowing for accurate resuspension to the required concentration.

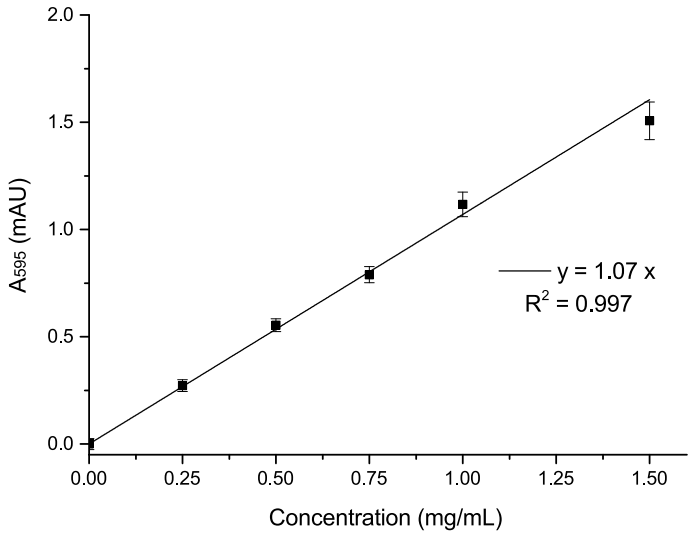


Figure 2.20. *A calibration curve was acquired using Bradford reagent and BSA as a standard. All samples were measured in triplicate and errors correspond to one standard deviation.*

2.4 Protein characterisation

2.4.1 Characterisation by mass spectrometry

Although the molecular weight of the proteins had previously been estimated by SDS-PAGE, the exact mass was yet to be determined. Liquid chromatography-mass spectrometry (LC/MS) is a common analytical tool used to isolate proteins, or molecules, from a sample or assay and measure their molecular weight.[101;169] High-performance liquid chromatography (HPLC) first separates the components of the protein sample depending on their retention time on a reverse phase HPLC column. Electrospray ionisation (ESI) mass spectrometry (MS) then ionises the sample and a magnetic or electric field sorts the ions into different charge states based on their mass-to-charge ratio (m/z). Each of these ions (I) are from a single intact protein with mass $M_r = n(I - N^+)$, where N^+ is the mass of a proton and n is a positive integer. The mass of the intact protein is then calculated through a maximum entropy deconvolution algorithm with defined mass range boundaries. In essence, this method maximises the entropy (H):

$$H = -K \sum_{i=1}^{n} p_i \log(p_i),$$

where K is a constant and p_i is the probability of a system being in cell i of its phase space.

By selecting the mass spectra corresponding to the retention time of the protein, and selecting a mass range in which the protein is expected to lie, a deconvoluted spectrum may be acquired.

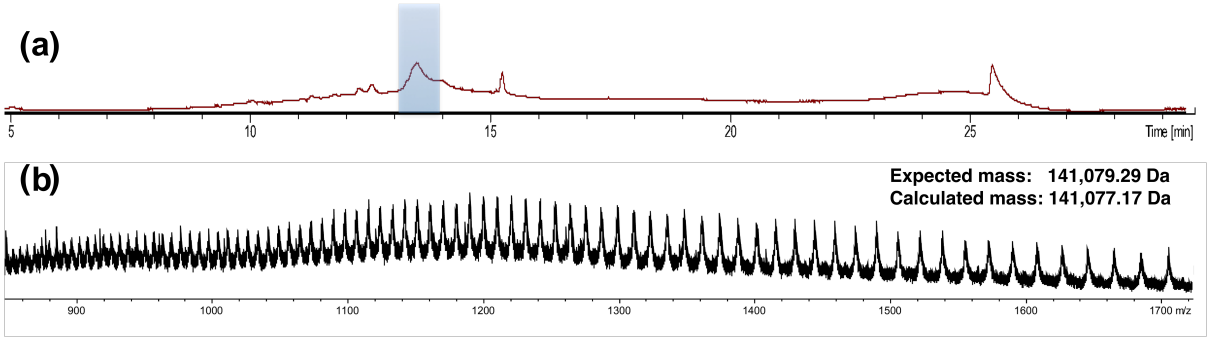


Figure 2.21. *UV chromatogram of QMN module 5* ***(a)*** *and the corresponding mass spectra at* $t = 13 - 14$ *minutes* ***(b)***.

LC/MS was used to confirm each protein had the expected molecular weight. The LC component allowed all purified proteins to be separated from contaminants. The peak with the retention time corresponding to the eluted protein was selected (Figure 2.21 (a)). All mass spectra of samples eluted during this time period were combined with the resulting spectra consisting of multiple peaks corresponding to different charge-states of the protein (Figure 2.21 (b)). The spectra were then deconvoluted using Compass DataAnalysis 4.1's (Bruker) maximum entropy deconvolution algorithm to provide the measured mass of each protein (Figure 2.22).

As can be seen in Table 2.2, the measured molecular weights were within 50 ppm for all purified proteins. As noted in the SDS-PAGE analysis, a second peak was observed with a higher molecular weight (22,107 Da) for M3_ACP. This peak had approximately 20 % of the intensity of the peak corresponding to the predicted molecular weight of M3_ACP. The nature of this species is discussed in section 2.4.3.

2.4.2 Characterisation by circular dichroism spectroscopy

Ultraviolet (UV)-circular dichroism (CD) is an analytical technique commonly used to analyse the secondary structure of proteins. It measures the differential absorption of left and right polarised light caused by passing through an optically active medium, such as the protein sample. Each secondary structure element will impart a distinct spectral

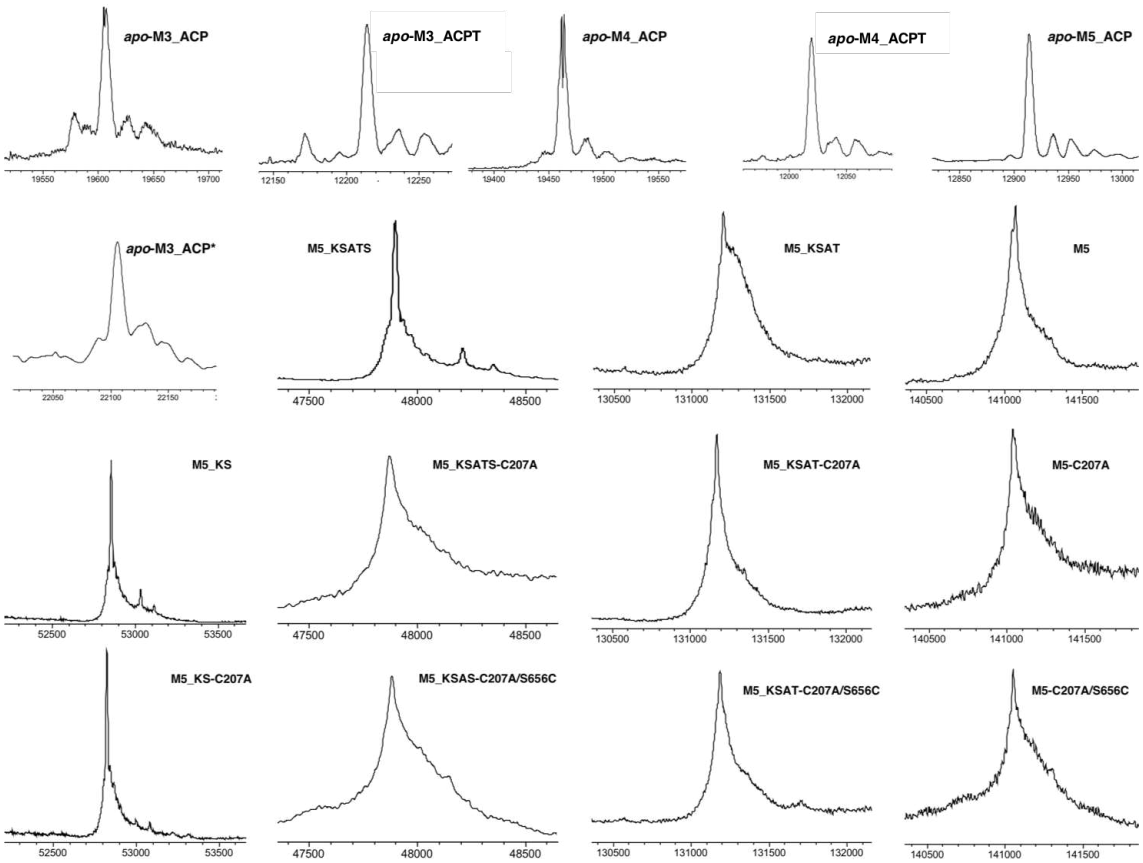


Figure 2.22. *Deconvoluted mass spectra for all overproduced proteins (from spectra in Figure 8.4). Expected and measured masses are displayed in Table 2.2.*

Table 2.2. *The measured and expected molecular weights of overproduced proteins.*

Protein domain	M3_ACP (*)	M3_ACPT	M4_ACP	M4_ACPT	M5_ACP	M5_KS	M5_KSATS	M5_KSAT
Expected mass (Da)	19,608.83 (unknown)	12,212.79	19,465.05	12,019.77	12,915.41	52,862.21	97,887.79	131,211.28
Measured mass (Da)	19,609.32 (22,105.90)	12,213.42	19,464.17	12,020.02	12,914.3875	52,861.13	97,886.15	131,209.80
Protein domain	**M5**	**M5_KS-C207A**	**M5_KSATS-C207A**	**M5_KSATS-C207A/S656C**	**M5_KSAT-C207A**	**M5_KSAT-C207A/S656C**	**M5-C207A**	**M5-C207A/S656C**
Expected mass (Da)	141,079.29	52,830.15	97,855.73	97,871.79	131,179.22	131,195.28	141,047.23	141,063.29
Measured mass (Da)	141,077.17	52,829.10	97,854.56	97,870.43	131,178.57	131,193.36	141,047.10	141,059.78

signature. For instance, α-helices have negative maxima at 208 and 222 nm, as well as a positive maximum at 190 nm. In contrast, β-strands display a negative maximum at 218 nm and a positive maximum at 195 nm.

Spectra were obtained for all wild type protein constructs (Figure 2.23) and Dichroweb was used to calculate the percentage of α-helix, β-strand and disordered region (Table 2.3) in the protein complex. Phyre2 was then used to predict the expected percentages of α-

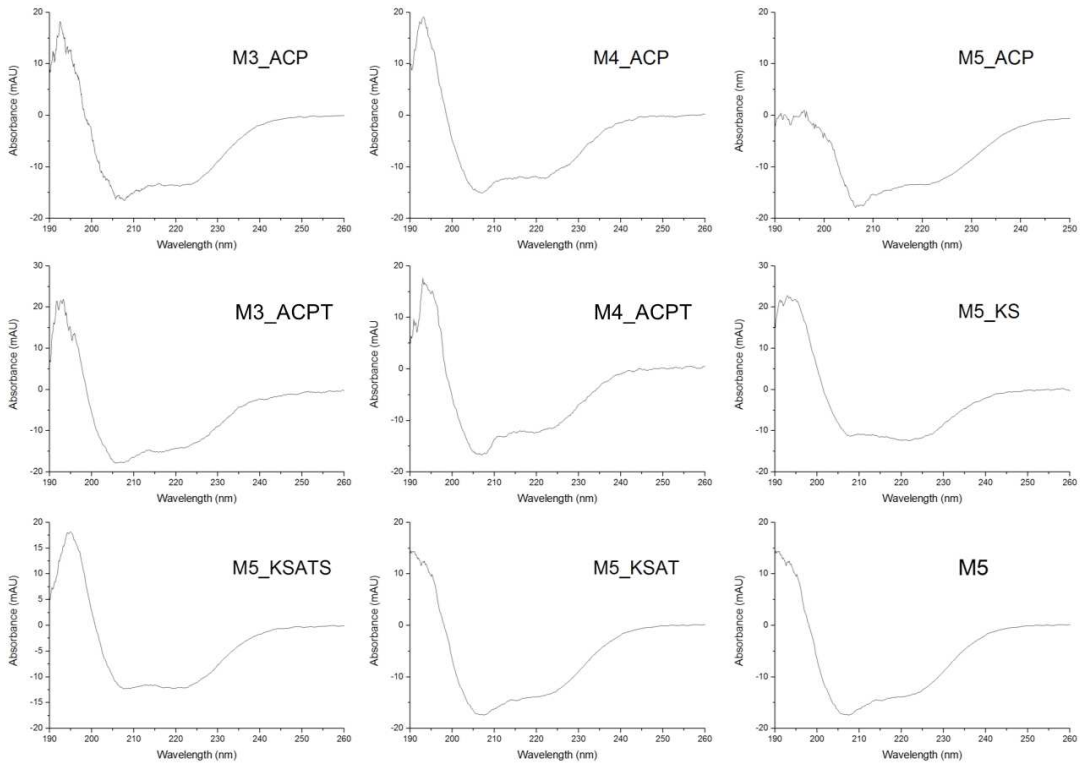


Figure 2.23. *CD spectra of purified wild type proteins from the QMN PKS pathway.*

helices, β-strand and disordered region of each protein/complex and comparisons were made between the calculated and expected values for each type of structure. The experimental and predicted values displayed contiguousness for the percentage of α-helix in each protein. For the ACPs, the percentage of β-strand predicted by Dichroweb was lower than for the experimental data. However, low percentages of β-strand can be predicted by Dichroweb for proteins with solved crystal structures showing they contain no β-strands (e.g. concanavalin). Therefore, the values obtained are indicative of correct folding of M3_ACP, M3_ACPT, M4_ACP, M4_ACPT, M5_ACP, M5_KS, M5_KSATS, M5_KSAT and M5.

2.4.3 Characterisation of the glycosylated M3_ACP

Other strains of *E. coli* were used to confirm the higher mass protein co-expressed with M3_ACP (M3_ACP*) was not just an artefact of BL21* overproduction (Figure 2.24). Three BL21* derivatives were used; Rosetta^{TM} 2 (DE3) singles, which is designed to enhance the expression of rare codons, C43, which is designed for the overproduction of toxic proteins and NiCo21, which is engineered to minimise the co-expression of endogenous *E.*

Table 2.3. *Measured (and predicted) secondary structures of purified wild type proteins based on Dichroweb predictions (and homology models).*

Protein domain	α-helix (%)	β-strand (%)	Disordered (%)
M3_ACP	60 (60)	3 (0)	26 (34)
M3_ACPT	69 (69)	3 (0)	12 (14)
M4_ACP	61 (61)	2 (0)	34 (32)
M4_ACPT	71 (69)	7 (0)	0 (14)
M5_ACP	66 (69)	7 (0)	14 (16)
M5_KS	40 (42)	18 (16)	24 (8)
M5_KSATS	43 (42)	12 (16)	25 (9)
M5_KSAT	34 (37)	13 (16)	26 (10)
M5	38 (41)	12 (13)	27 (10)

coli metal-binding proteins, therefore allowing for superior IMAC purification.

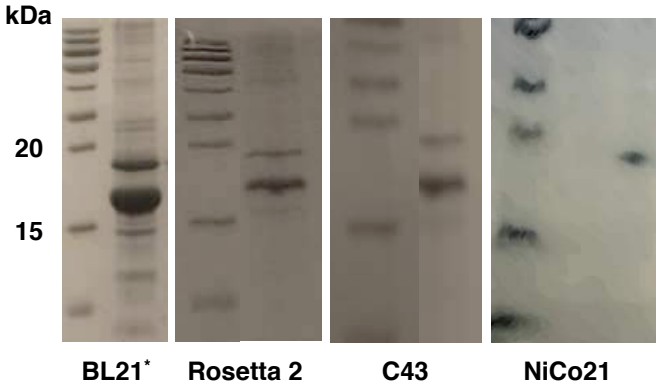


Figure 2.24. *SDS-PAGE analysis following IMAC M3_ACP overproduction in different E. coli (DE3) strains.*

Visualisation by SDS-PAGE gel electrophoresis indicated that the higher mass band was still present for each sample, with the exception of protein overproduced in NiCo21. However, MALDI (matrix-assisted laser desorption ionisation) analysis of the protein sample showed the protein with higher molecular weight was still present (Figure 2.25).

A range of chromatographic techniques were utilised in an attempt to separate the proteins (Figure 2.26). Neither a gradient of increasing imidazole nor the use of cobalt in place of nickel (due to its higher selectivity towards His-tagged proteins)[170] resulted in separation. Anion exchange chromatography and gel filtration were then attempted, again

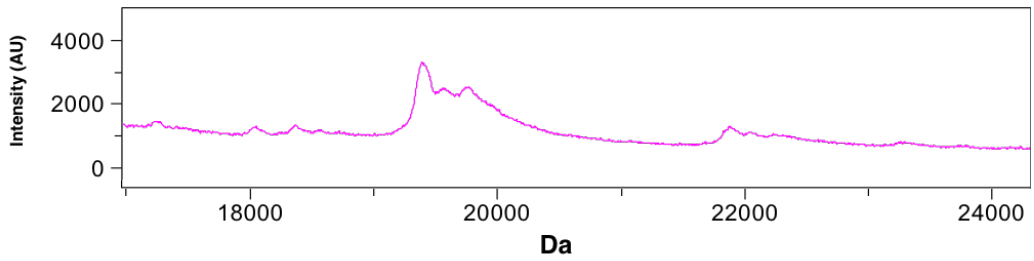


Figure 2.25. *MALDI analysis of M3_ACP confirmed the presence of two proteins with mass of approximately 19.5 and 22 kDa.*

resulting in both proteins being eluted together.

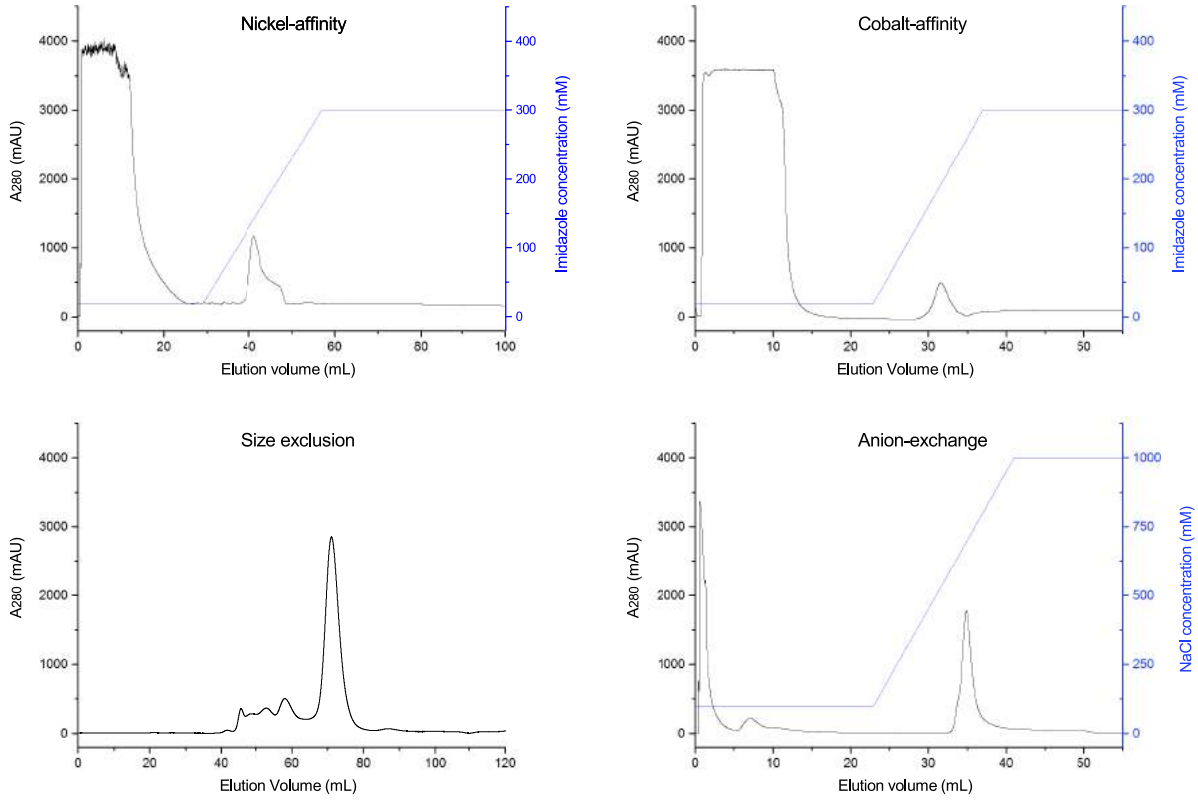


Figure 2.26. *Chromatograms of attempts to separate M3_ACP from the protein with higher molecular weight.*

The continued co-elution of the proteins suggested interactions were occurring between them, possibly because the protein with higher molecular weight is a posttranslationally modified (PTM) variant of M3_ACP.

One possibility was that the His-tag was being glycosylated. Cleavage of the His-tag resulted in bands corresponding to cleaved M3_ACP and a higher mass band, either cor-

responding to uncleaved M3_ACP or the higher mass PTM-ACP after cleavage. Complete disappearance of the higher mass His-tagged PTM-ACP suggests it is likely to be the latter. Therefore, it is probable that the PTM does not occur on the His-tag.

Peptide mass fingerprinting (PMF) was used to see if both proteins contained sequence similarity in order to confirm that the higher mass band corresponds to a PTM M3_ACP. PMF is a technique in which a protein is enzymatically digested, usually by trypsin, before the resulting peptides' masses are measured using MS. Therefore, if the higher mass protein is a PTM M3_ACP then it would likely yield a number of peptides with the same mass as M3_ACP after such treatment, confirming that both proteins share a sequence similarity.

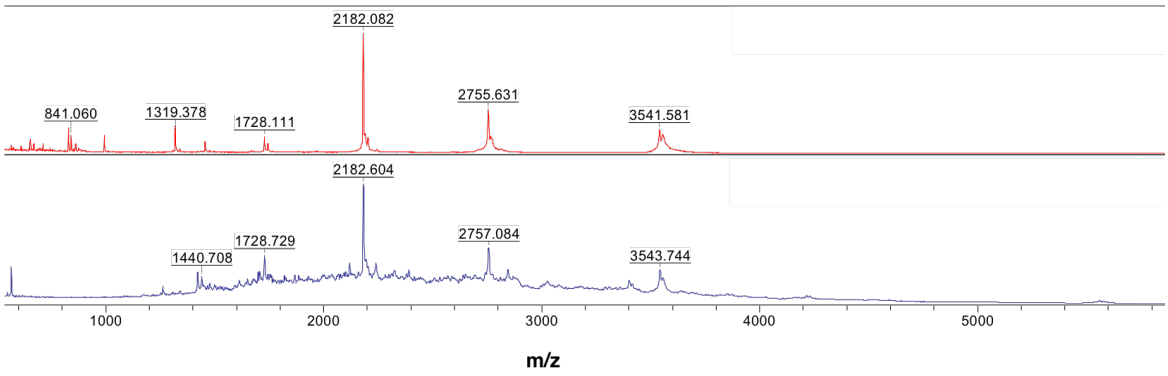


Figure 2.27. *MALDI analysis of peptides resulting from tryptic digests of M3_ACP (top) and a proposed posttranslationally modified M3_ACP (bottom).*

The protein mixture was analysed by SDS-PAGE and both proteins' bands were excised separately from the gel. Tryptic digests were carried out on both gel pieces and the resulting peptides were analysed using MALDI (Figure 2.27). The expected peptide fragments were observed for both proteins, with 82 % sequence coverage for M3_ACP and a 68 % sequence coverage for the protein with the higher molecular weight. This shows that the higher molecular weight protein has high sequence similarity to M3_ACP and is likely to be a posttranslationally modified M3_ACP.

The modification was proposed to be an N-linked glycosylation based on the identification of an NGT motif within M3_ACP. An N-linked glycosylation is the attachment of an oligosaccharide, known as a glycan, to the amine group of an asparagine residue, resulting in the production of a glycoprotein. 72.2 % of reported N-linked glycosylations occur at the asparagine of an NxT motif, with NGT being the most common motif (12.6 %).[171]

Furthermore, the location of the NGT motif in the docking domain region would explain why M3_ACPT did not contain the second protein. A peptide with an increased mass of 2.5 kDa could not be seen in the MALDI spectra for the digested higher molecular weight protein, perhaps due to glycopeptides not ionising completely due to the microheterogeneity of attached glycans.

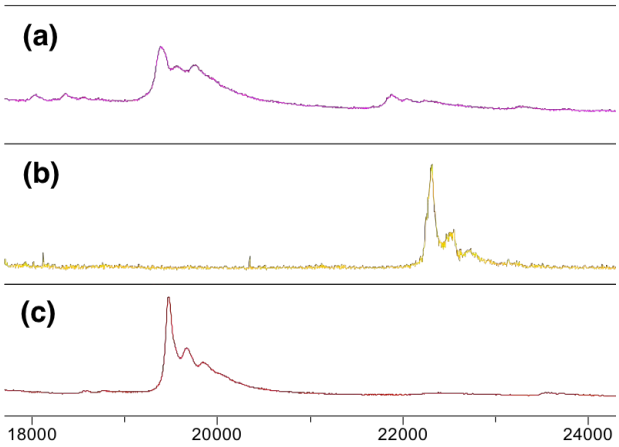


Figure 2.28. *MALDI analysis of **(a)** M3_ACP, **(b)** the isolated higher mass band (22 kDa) and **(c)** the isolated higher mass band after deglycosylation (19.6 kDa).*

In order to confirm this hypothesis, and attempt to remove the potential mutation, an in-gel deglycosylation kit was used. The kit allows peptides to be extracted from SDS-PAGE gel pieces and for an amidase (PNGase F) to cleave all N-linked oligosaccharides. The successful N-deglycosylation of the higher molecular weight protein to a protein of the expected molecular weight of M3_ACP was observed (Figure 2.28), thereby confirming the higher molecular weight protein to be a N-linked glycosylation of M3_ACP.

2.5 Conclusions

M3_ACP, M3_ACPT, M4_ACP, M4_ACPT, M5_ACP, M5_KS, M5_KSATS, M5_KSAT and M5 were recombinantly overproduced in a soluble form. Each construct contained an N-terminal His-tag, enabling purification by IMAC. Mutations of the KS and AT active-sites were also introduced in M5_KSATS, M5_KSAT and M5, allowing for the probing of intramodular AT-ACP interactions as well as providing controls. In addition, M5_KS-C207A was produced to provide a control for M5_KS. M5_KSATS, M5_KSAT and M5 were

further purified by gel filtration to enable high-resolution structural studies. Comparison to a range of protein markers showed a loss of dimerisation between M5_KSATS (residues 1 - 888) and M5_KSAT (residues 1 - 1198), suggesting the structure of the AT-ACP linker region must play a vital role in module dimerisation.

The purity of all proteins was checked by SDS-PAGE and the molecular weight characterised by mass spectrometry before the correct secondary structure was confirmed using circular dichroism spectroscopy. Nanodrop estimates of protein concentration did not agree with estimates from Bradford assays for M3_ACPT, M4_ACP and M4_ACPT, perhaps due to the lack of residues with absorbance at 280 nm. To ensure accurate concentrations, all ACPs were lyophilised, weighed and resuspended.

In addition, IMAC purified M3_ACP was shown to contain a second protein with an increased molecular weight of approximately 2.5 kDa. The larger protein was not present in the truncated M3_ACPT and the two proteins could not be separated by a variety of chromatographic techniques. PMF analysis confirmed the higher mass protein was likely to be a PTM modified version of M3_ACP. The PTM was predicted to be an N-glycosylation due to the presence of an NxT motif in M3_ACP's docking domain. This was confirmed by the successful deglycosylation of the higher mass protein to regenerate a protein with the expected molecular weight of M3_ACP.

Chapter 3

The intramodular interactions of quartromicin module 5

3.1 Confirming the catalytic activity of module 5

M3_ACP, M3_ACPS, M4_ACP, M4_ACPS, M5_ACP, M5_KS, M5_KSATS, M5_KSAT and M5 were all previously overproduced and purified (Table 2.1), along with KS active-site mutants (C207A) and KS/AT active-site mutants (C207A/S656C).

The proposed role of M5 in the PKS pathway is to sequester an 8-carbon substrate (from M3_ACP) or a 10-carbon substrate (from M4_ACP) and to utilise a malonyl-CoA building block to catalyse chain extension by two carbons.[155] A series of *in vitro* bioassays were planned to confirm the intramodular catalytic activity of M5 and also to explore the module's ability to utilise different extender units and substrates.

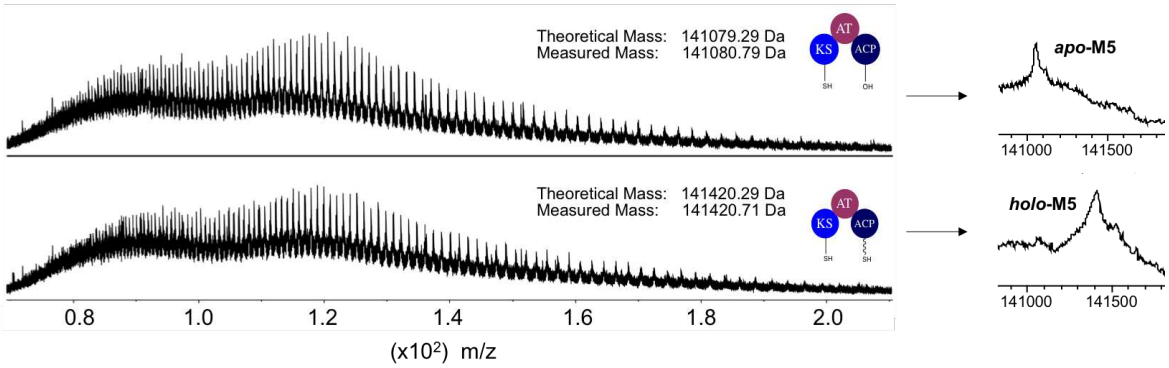


Figure 3.1. *A LC/MS-QTOF chromatogram of apo-M5 (top left) and a chromatogram confirming the successful phosphopantetheinylation of M5 (bottom left). Deconvoluted mass spectra corresponding to the intact proteins are also shown (right).*

An extension assay[172] was then designed to confirm the ability of the KS to sequester a pantetheine-bound polyketide, the AT to utilise a malonyl-derived extender unit and the modules' ability to successfully catalyse the translocation of an extender unit to M5_ACP and subsequent extension of the KS-bound polyketide by incubation with intact M5, a CoA-bound extender unit and a pantetheine-bound substrate mumic. Successful conversion from *apo-* to *holo*-M5 was confirmed by LC/MS (Figure 3.1). *Holo*-M5 was then incubated with malonyl-CoA and octanoyl-SNAC. Octanoyl-SNAC was chosen as a mimic of the polyketide transferred from M3_ACP to M5_KS. Malonyl-CoA was chosen as it is the predicted extender unit utilised by M5. LC/MS analysis of the assay mixture

showed a peak in the deconvoluted mass spectra at 141,588 Da, corresponding to 3-keto-decanoyl-*holo*-M5. Collision-induced dissociation (CID) provided confirmation the change had occurred on the pantetheine moiety, due to the presence of ejected 3-keto-decanoyl-pantetheine (429.24 Da, Figure 3.2).

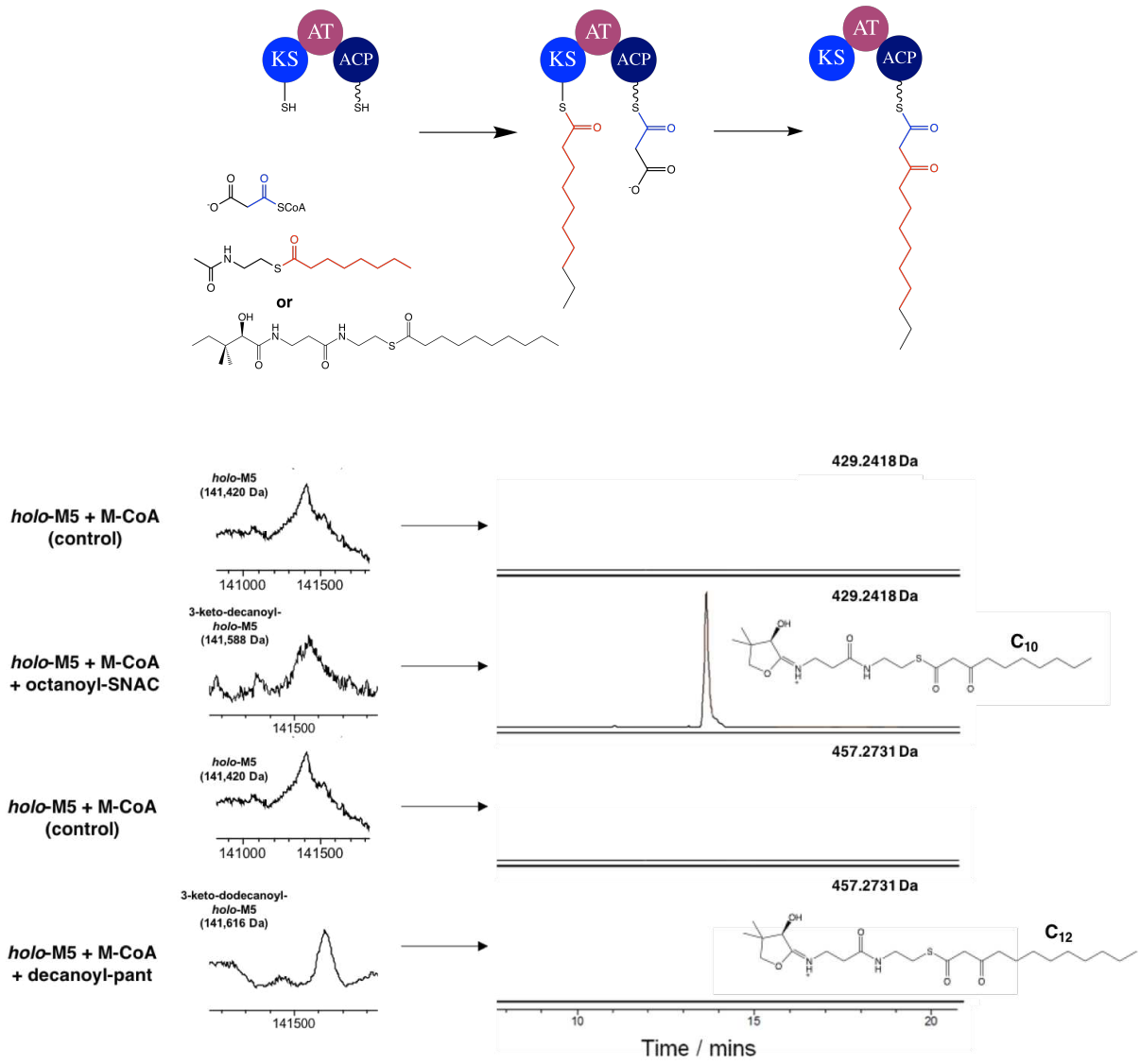


Figure 3.2. *Schematic representation of an SNAC-bound polyketide subunit being extended by a malonyl extender unit, resulting in the extension of the polyketide chain by one keto group (top). Deconvoluted spectra (bottom left) and corresponding EICs (bottom right) from LC/MS analyses of holo-M5 incubated with octanoyl-SNAC or decanoyl-pantetheine. Assay components are shown on the left of each EIC. Ejected pantetheines and m/z are displayed on the right (± 0.05 Da).*

CID results in a proportion of the phosphopantetheinyl moiety being ejected from *holo*-ACPs, leaving a phosphate ion attached to the ACP's PTM serine residue. In single domain ACPs, this dissociation has to be forced by supplying energy, but due to the large molecular weight of M5, it occurs spontaneously during ionisation. A five-membered tetrahydrofuranol ring with a protonated imine is formed when the carbonyl of an amide

displaces the pantetheinyl unit (Figure 3.3). The simplicity of detecting relatively small fragment ions through extracted ion chromatograms (EIC), compared to the mass change of an intact protein, is particularly practical in analysing whether a large protein complex, such as M5, has successfully extended a polyketide substrate.

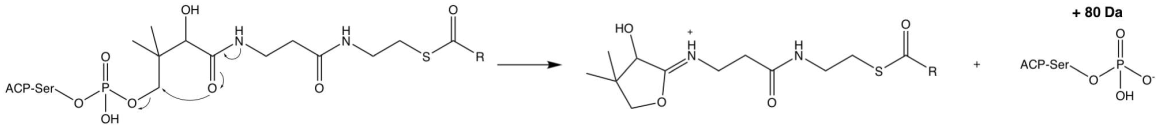


Figure 3.3. *The mechanism of CID induced pantetheine ejection from a holo-ACP.*

The experiment was then repeated with decanoyl-pantetheine, in place of octanoyl-SNAC, as a mimic of the substrate transferred from M4_ACP. A pantetheine substrate was used because longer chain SNAC substrates are likely to be insoluble. Although deconvolution of the spectra resulted in a peak corresponding to 3-keto-dodecanoyl-*holo*-M5 (141,616 Da), no ejected 3-keto-dodecanoyl-pantetheine (457.27 Da) was detected.

Combined, these experiments confirm the ability of the AT domain of module 5 to load malonyl-CoA and for the KS domain to accept either 8-carbon or 10-carbon chains and catalyse chain extension, fulfilling the predicted role of M5 in the biosynthetic pathway.

3.1.1 Exploring module 5's ability to utilise different malonyl-derived extender units

To produce the predicted polyketides necessary to form the skeleton of quartromicin, M5 is proposed to only use malonyl-CoA extender units. Furthermore, the AT domain contains a conserved HAFH motif in the active-site which is proposed to control the module's extender unit specificity, meaning it can only catalyse the sequestering of malonyl extender units.[173] ATs with a conserved YASH motif are reported to utilise methylmalonyl extender units. This is proposed to be because the limited room in the binding pocket of ATs containing a HAFH motif, due to the extra molecules of the phenylalanine residue (as opposed to a serine), prevents the methylmalonyl extender unit from entering the binding pocket. This would prevent the AT domain from catalysing the sequestering of a

methylmalonyl extender unit and result in no extended product.

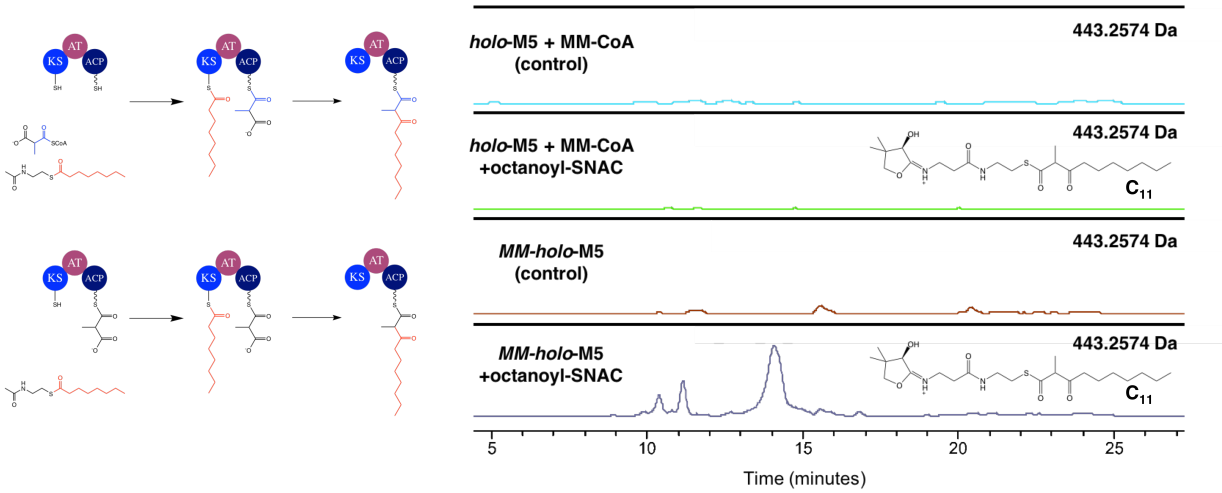


Figure 3.4. *Schematic representations of M5 extending an octanoyl-SNAC with a methylmalonyl extender unit (left) and EICs from LC/MS analyses following incubation of octanoyl-SNAC with holo-M5 and methylmalonyl-CoA (top) and methylmalonyl-holo-M5 (bottom). Components of the assays are displayed on the left. Ejected pantetheines and m/z are displayed on the right (± 0.05 Da).*

Holo-M5 was incubated with octanoyl-SNAC and methylmalonyl-CoA. LC/MS analysis showed no peaks corresponding to ejected pantetheine containing an extended product (Figure 3.4). This suggests that M5 is only able to utilise a malonyl extender unit due to the AT's inability to catalyse the sequestering of a methylmalonyl extender unit for the ACP. To confirm whether an elongated product could be produced if a methylmalonyl extender unit was loaded directly onto the ACP, M5 was directly converted to methylmalonyl-*holo*-M5 by incubation with methylmalonyl-CoA, Sfp and $MgCl_2$, before incubation with octanoyl-SNAC. A peak was observed with m/z 443.26 Da, which was absent in the control, corresponding to the successful extension of the octanoyl substrate with a methylmalonyl extender unit. This indicates that the extender unit selectivity of M5 is due to the AT domain's inability to sequester a methylmalonyl extender unit.

3.1.2 Systematically probing the substrate specificity of module 5

The second consideration in the intramodular functionality of M5 was which substrates M5 could sequester and elongate with a malonyl-extender unit. PKS modules have previously been shown to have a wide substrate tolerance, with a range of tested polyketide

substrates successfully elongated.[174] Confirming the substrate tolerance of M5 will help elucidate whether it plays a role in ensuring the PKS pathway's fidelity.

A number of SNAC/pantetheine-bound substrates were synthesised by Matthew Beech, Christian Hobson, Dr. Joleen Masschelein and Dr. Shan Shan Zhou (Figure 3.5).[175] Whilst the shorter chain substrates were synthesised as SNAC derivatives, the dodecanoyl substrate was coupled to pantetheine. This was due to potential solubility issues when coupled to SNACs, as discussed for decanoyl-pantetheine previously.

A number of substrates were used to determine whether the length of the carbon chain has any effect on substrate tolerance. The previous elongation assays were repeated with propionyl-SNAC (3-carbon chain) and dodecanoyl-pantetheine (12-carbon chain). Successful elongation was seen in both cases (Figure 3.6, 359.19 and 485.30 Da, respectively), suggesting substrate length has no role in selectivity. Next, a *tert*-butanoyl-SNAC was utilised to test whether additional methyl groups would have an effect on substrate tolerance, resulting in detection of extended product (387.19 Da). Unsaturated-hexanoyl was also tested to determine whether double bonds in the carbon chain had an effect, with extended product detected (397.18 Da). Furthermore, 3-keto-octanoyl-SNAC was also successfully loaded and extended (443.22 Da), although in significantly reduced quantities (< 5 %). This potentially suggests that M5 less efficiently extends polyketides with a 3-keto group, although the reduction in detected product could potentially be due to an increased level of hydrolysis of the β-keto thioester.

Studying substrate selectivity has shown that M5 tolerates a wide range of substrates. Therefore, it is unlikely that M5's substrate tolerance plays a significant role in maintaining the selective promiscuity of the QMN PKS module skipping mechanism.

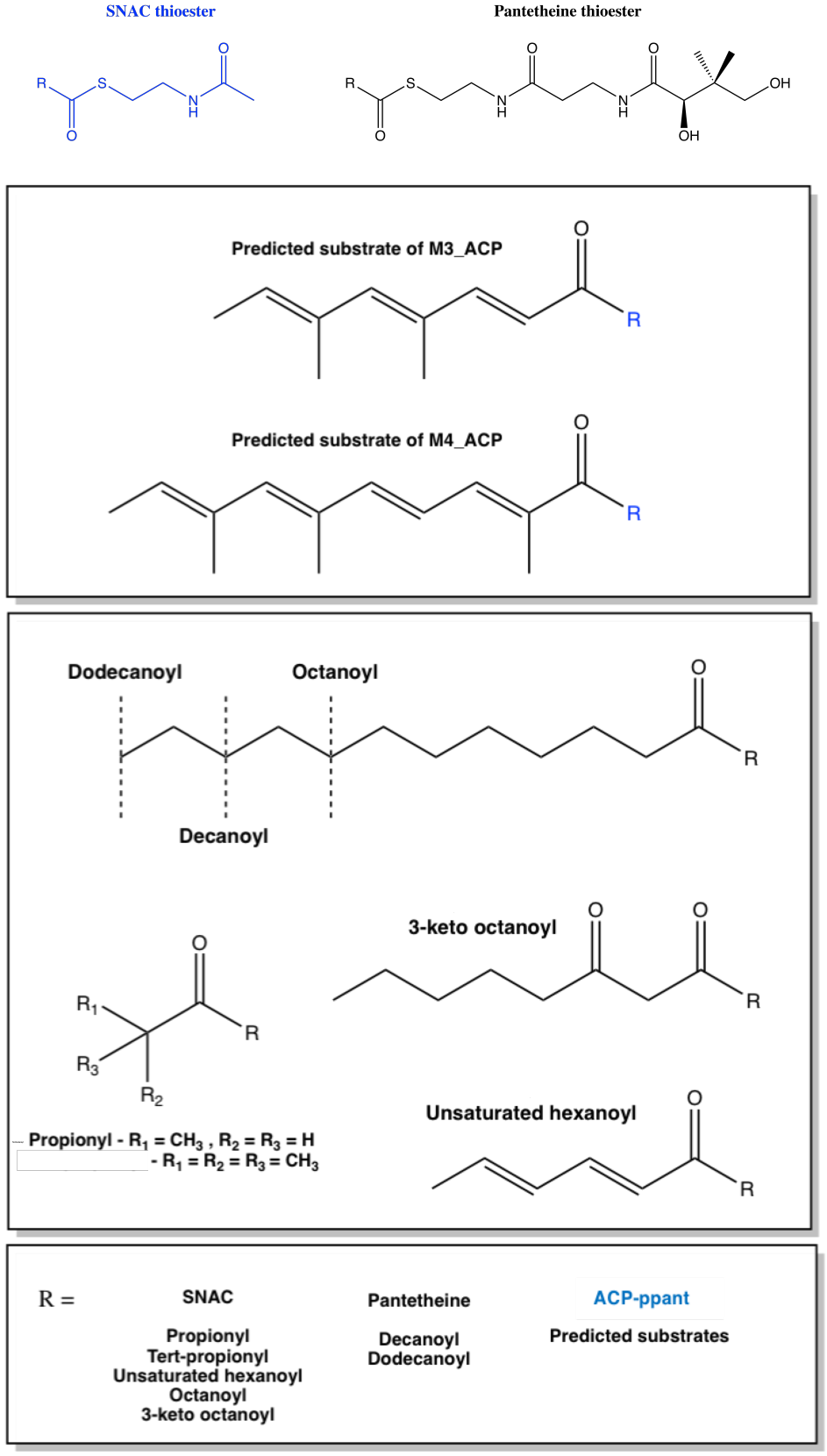


Figure 3.5. *S-N-acetylcysteamine (top left) and pantetheine (top right) thioesters.The SNAC and pantetheine-bound substrates used to probe M5's substrate tolerance (bottom).*

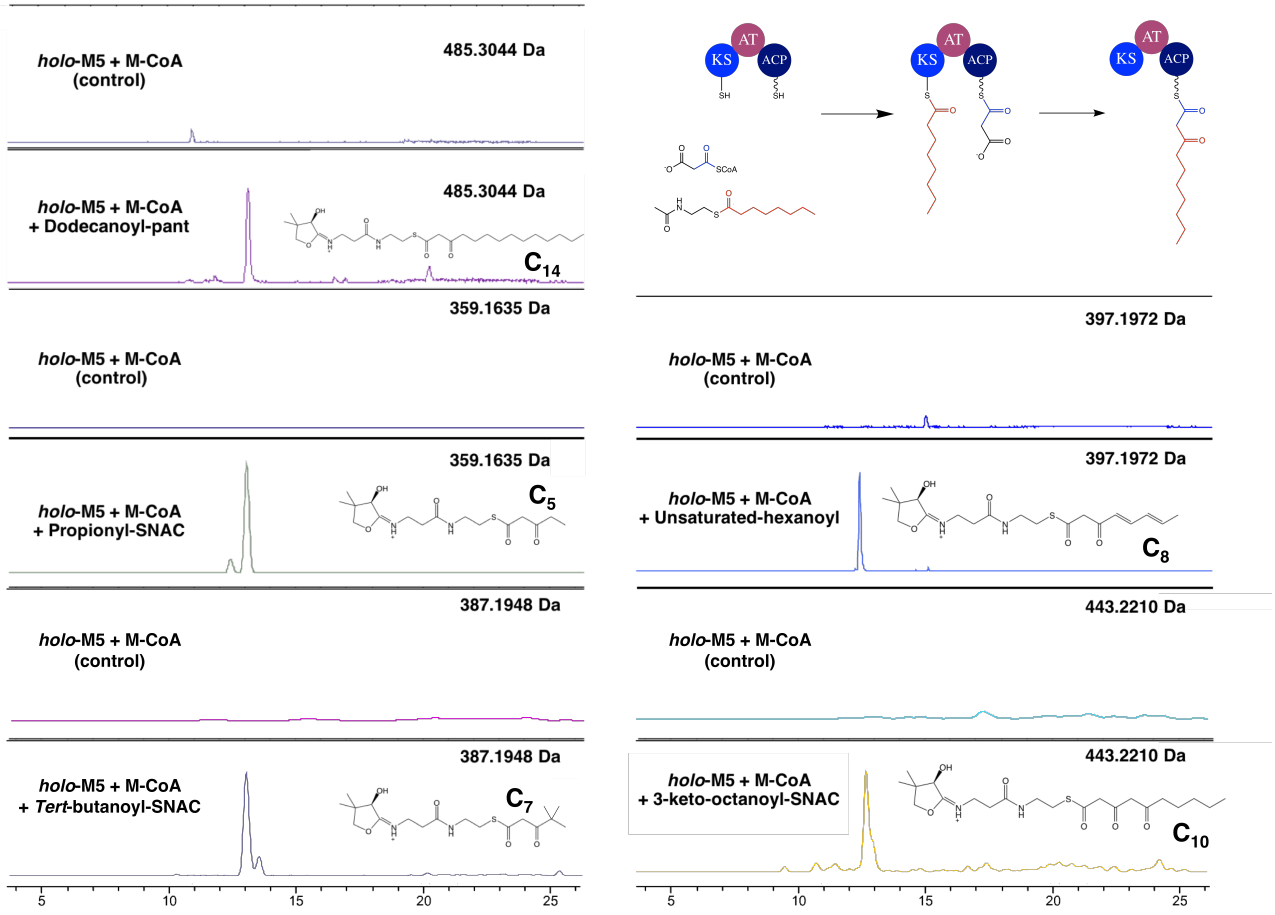


Figure 3.6. *EICs for LC/MS analyses following incubation of holo-M5 with malonyl-CoA and various SNAC/pantetheine-bound substrates. Components of the assays are displayed on the left. The ejected pantetheines and m/z are displayed on the right (± 0.05 Da).*

3.2 Intramodular cross-linking of quartromicin module 5

The ability to structurally characterise protein-protein interactions within PKS systems is severely limited due to their transient nature. In order to study such interactions, a site-specific enzymatic cross-linking approach may be utilised.[176] A pantetheine cross-linker (Matthew Jenner, pantetheine chloroacrylamide), containing a terminal chlorine leaving group, may be loaded onto an ACP's PTM serine through incubation with CoA synthase enzymes. The ACP-loaded linker should then be delivered to the active-site of a compatible KS domain, where the thiol of the active-site cysteine may attack the terminal chlorine of the cross-linker. This would result in the formation of a stable covalent linker between the KS and ACP domains of approximately the same length as a Ppant arm, en-

suring the KS and ACP domains are proximal to their positions during the elongation of a polyketide product. In addition to being a useful structural biology tool, the extent of cross-linking between two domains has been reported to correspond to the strength of protein-protein interactions occurring.[177] Therefore, not only does enzymatic cross-linking allow the presence of interactions to be identified, it also allows the estimation of affinity between two proteins.

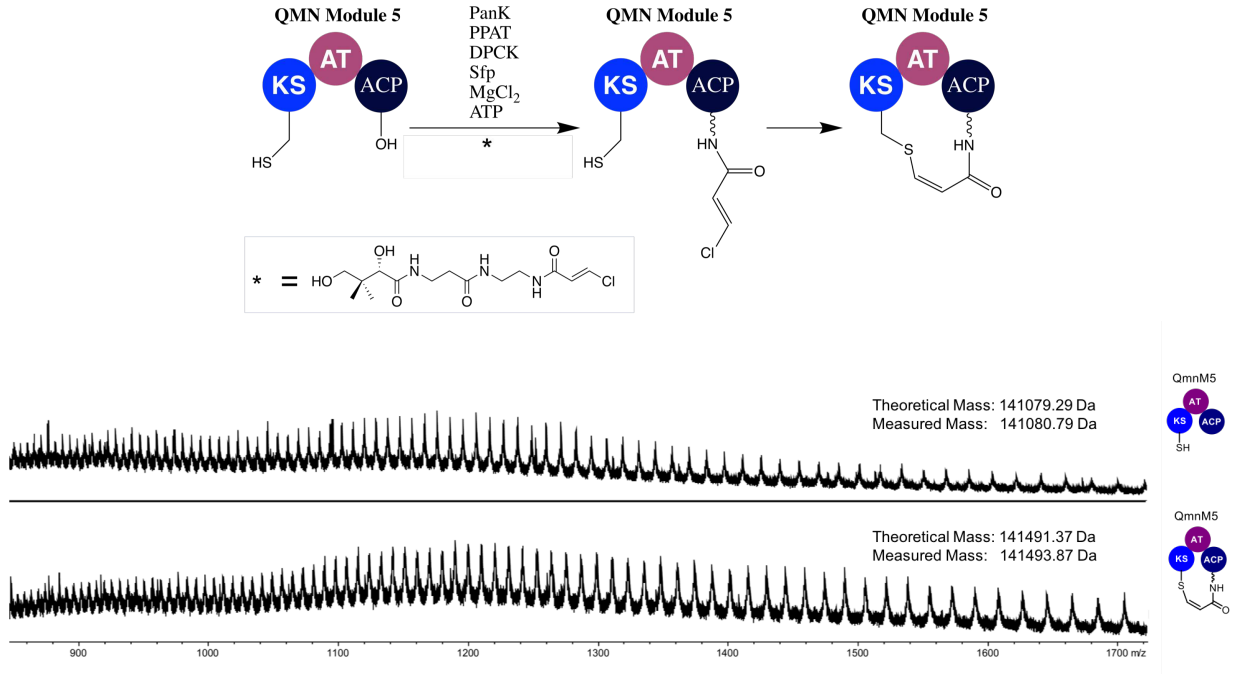


Figure 3.7. *Schematic representation of the intramodular cross-linking between the KS and ACP of M5 (top). LC/MS chromatograms of apo-M5 and intramodularly KS-ACP cross-linked M5 (KS-ACP, bottom).*

M5 was incubated with the cross-linker and the enzymes required for loading the cross-linker to the ACP's PTM serine, resulting in successful KS-ACP cross-linking as determined by an increase in molecular weight of 413 Da in the deconvoluted mass spectra of M5 (Figure 3.7).

M5 contains KS, AT and ACP domains, as well as a 310 residue linker region between the AT and ACP (AT-ACP linker). To determine the importance of each domain and the linker region, M5's ACP (M5_ACP) was incubated with pantetheine chloroacrylamide and either M5_KS, M5_KSATS (without linker region) or M5_KSAT (with linker region). Analysis by SDS-PAGE (Figure 3.8) showed successful cross-linking to the KS domain of M5_KSAT as a new band was present at 140 kDa, compared to the control in which no cross-linker was added.[178] Neither of the constructs lacking the AT-ACP linker (M5_KS and M5_KSATS) showed any sign of cross-linking.

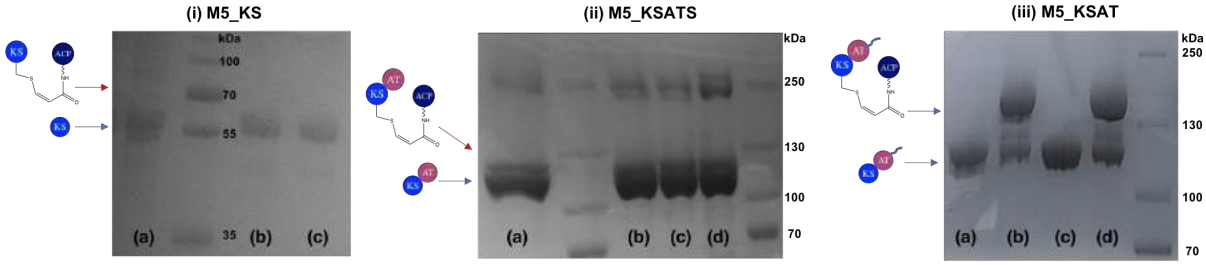


Figure 3.8. *SDS-PAGE analysis following attempted cross-linking of M5_ACP with **(i)** M5_KS, **(ii)** M5_KSATS and **(iii)** M5_KSAT variants. Variants in each lane are **(a)** WT control (in which cross-linker was omitted), **(b)** WT, **(c)** C207A mutants and **(d)** C207A/S656C mutants.*

C207A variants of each construct had previously been produced to prevent cross-linking to the KS active-site. These constructs acted as negative controls in order to show the specificity of the cross-linking interactions was to the KS domain. SDS-PAGE analysis of each C207A variant incubated with cross-linker loaded M5_ACP displayed no bands of increased mass, indicating no cross-linking had occurred and confirming the increased mass band seen for M5_KSAT corresponds to KS-ACP cross-linking.

C207A/S656C variants were also available where the serine of the AT active-site had been mutated to a cysteine. This would allow cross-linking between the AT and ACP domains, as well as ensuring no KS-ACP cross-linking occurred. SDS-PAGE analysis of M5_KSAT-C207A/S656C with cross-linker loaded M5_ACP resulted in a band of $\sim$ 140 kDa, confirming AT-ACP cross-linking. Constructs omitting the AT-ACP linker region (M5_KS-C207A/S656C and M5_KSATS-C207A/S656C) showed no sign of cross-linking.

3.3 Conclusion

The ability of M5 to sequester and elongate different extender units and polyketide substrates was studied. An 8-carbon substrate mimic and malonyl-extender unit successfully produced an elongated product, confirming the catalytic functionality of the construct. Probing of extender unit specificity indicated that M5 could only utilise a malonyl extender unit. M5 was able to extend a polyketide chain when a methylmalonyl extender unit was loaded directly onto the ACP, suggesting the module's extender unit specificity was due to the AT being unable to catalyse the sequestering of a methylmalonyl extender unit. This agrees with previous experimental data that suggests modules containing a

conserved HAFH motif are only able to utilise malonyl extender units.

M5 was shown to have a wide substrate tolerance, with elongated products detected for a number of different substrates. Varying lengths of substrate were successfully extended (3, 8, 10 and 12-carbons), as well as saturated chains and chains containing extra methyl groups and a 3-keto group. This suggests that the modules' substrate tolerance is extremely wide and that the fidelity of the pathway is likely governed by other factors.

The next step is to probe M3 and M4's intramodular catalytic functionality through the range of substrates and extender unit assays. Cloning and overproduction of constructs containing various regions of M3 and M4 would allow further cross-linking assays to determine the importance of each region in both modules' intramodular functionality. This will be discussed in chapter 5.

Interactions between M5_ACP and both the KS and AT domain of M5 were shown via chemical cross-linking, with no preference given for a specific domain. The ability to cross-link KS and AT domains to ACPs could be of significant advantage in structurally studying polyketide modules and domain complexes. Being able to ensure the proximity of KS or AT active-sites to the ACP's PTM serine could allow the study of transient conformational states, as well as potentially reducing structural mobility. This would allow techniques such as X-ray crystallography to be considered, possibly providing greater insights into the PKS system. The locking of M5 in both intramodular conformations, as well as with upstream and downstream modules, would also be potentially beneficial in obtaining cryo-EM maps of the module in each stage of its catalytic cycle.

KS-ACP and AT-ACP interactions were only observable for constructs containing the AT-ACP linker region, suggesting the region plays a vital role in ensuring the module's functionality. The importance of the AT-ACP linker region will be explored in chapter 4.

Chapter 4

Structural studies on module 5 of quartromicin

4.1 Introduction

Constructs containing various regions of M5 had previously been overproduced and purified (chapter 2, Figure 4.1).

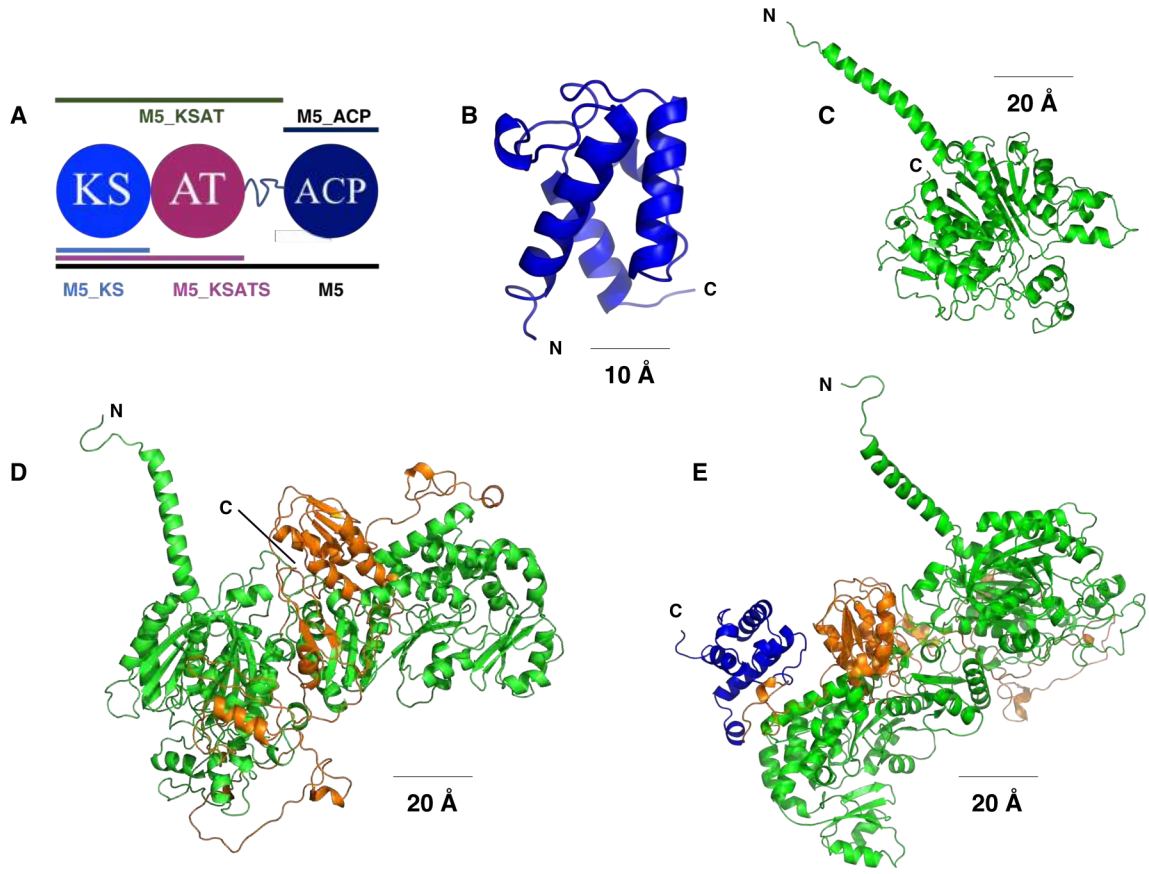


Figure 4.1. *Available constructs containing various domains and regions of M5* ***(A)*** *and homology models for M5_ACP* ***(B)****, M5_KS* ***(C)****, M5_KSAT* ***(D)*** *and M5* ***(E)****.*

There are various ways of obtaining structural information on a biological system. The most basic is gel filtration, in which proteins are separated according to size, allowing the determination of the protein's oligomerisation state. To obtain more detailed information about the molecular interactions within proteins, a variety of methods may be used depending on the size and flexibility of the construct of interest.

Relatively small proteins (< 20 kDa) may be structurally elucidated by NMR. The fact that over 10,000 NMR structures can be found on the PDB (including numerous ACP domains) attests to the technique's capabilities, with protein complexes containing over

1,000 residues being structurally elucidated by solution-state NMR.[179–181;55;182] Other domains that are too large to be solved by NMR, but less structurally mobile, may potentially be crystallised, opening up the possibility of utilising X-ray crystallography. Domains which have been structurally solved by X-ray crystallography include KR,[110;111] DH,[112;113] ER[33] and TE domains,[114;115] as well as KS-AT[108;116;117] and KR-ER didomains,[76] commonly to resolutions of below 2 Å.

Entire modules or complexes that are too structurally mobile to crystallise may be structurally elucidated by cryogenic-electron microscopy (cryo-EM).[117;123] However, the reduced resolution (5 - 10 Å) requires modelling in of high-resolution homologous structures for each domain to acquire information on a residue level.

4.2 Structural elucidation of module 5's ACP domain

Circular dichroism (CD) was used to determine conditions in which M5_ACP was well folded. Variable pH-CD indicated that M5_ACP (pI = 5.08) was structurally stable between pH 5.5 and 8.0, with a reduction in absorbance at 208 nm (indicative of unfolding) observed only for pH 5.0 (Figure 4.2). Variable-temperature CD showed a reduction in absorbance at 208 nm between 35 and 40^oC.

[U-^{15}N]-M5_ACP was overproduced (using minimal media containing $^{15}N\,H_4\,Cl$) and purified, and HSQCs were acquired in phosphate buffer. A range of different temperatures and pH values in which the protein had previously been shown to be stable were investigated (Figure 4.3). The HSQC's showed a wider dispersion of peaks at pH 6.5 than pH 7.0, and for 15^oC compared to higher temperatures. This is indicative of higher levels of folding and would likely provide spectra that are easier to assign. Promisingly, the sample retained structural integrity for 24 hours, allowing a larger number of scans to be undertaken. However, due to the number of solved ACP structures and high similarity between ACPs, other structural work was prioritised until a cryo-EM map of the entire module could be elucidated.

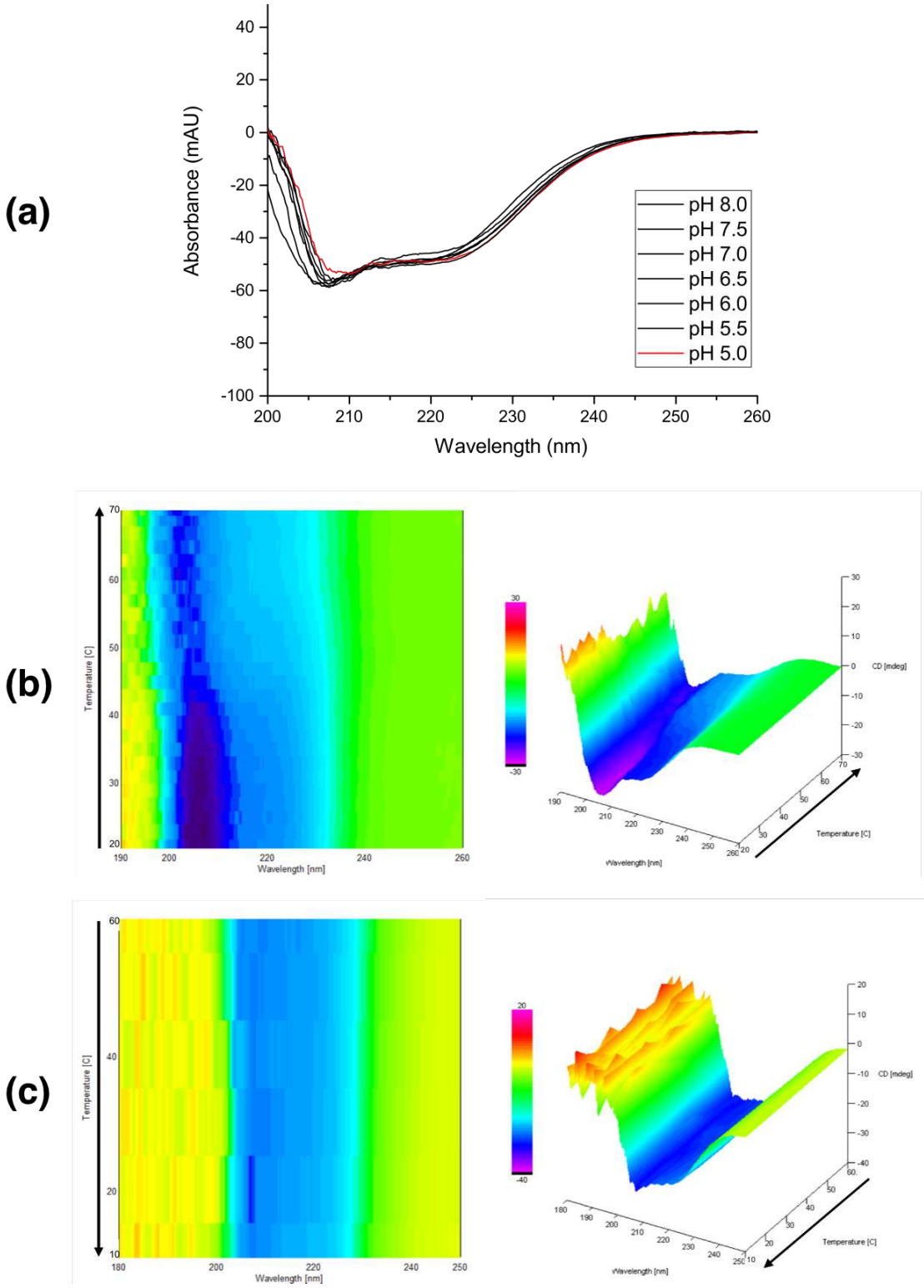


Figure 4.2. *Variable-pH **(a)** and temperature CD spectra for M5_ACP. The variable-temperature correspond to raising **(b)** and reducing temperatures **(c)**.*

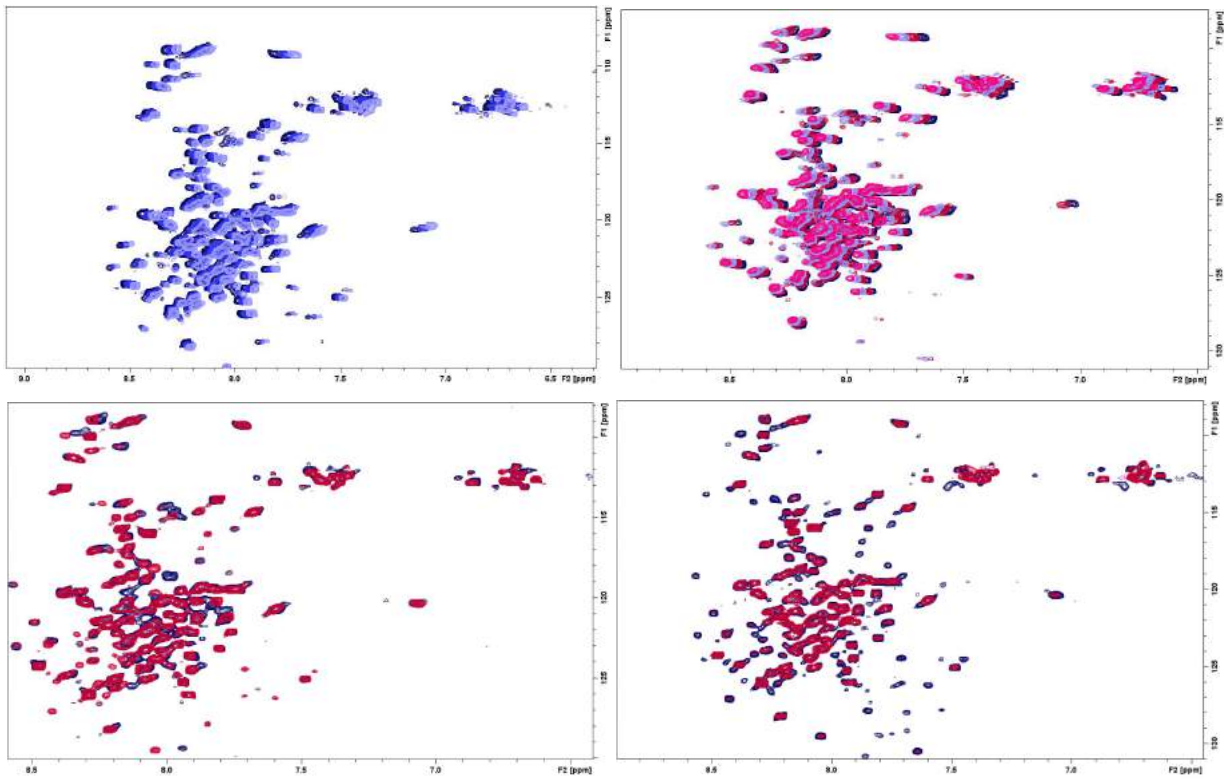


Figure 4.3. *HSQC spectra of [U-^{15}N]-M5_ACP. pH 6.5 (top left) at 15, 20 and 25 °C, and pH 7.0 (top right) at 10, 15, 20 and 25 °C. An HSQC of M5_ACP obtained at 15 °C (bottom left) displays pH 6.5 and pH 7.0 and pH 6.5 (bottom right) on the day the sample was prepared and a day later at 15 °C. NS = 32 (top left, top right and bottom left) and 132 (bottom right).*

4.3 Structural elucidation of quartromicin module 5's KS-AT didomain

As discussed in section 4.1, crystallisation of intact M5 would likely prove challenging due to the presence of the structurally mobile ACP domain and its large molecular weight. Initial trials were therefore carried out with M5 _KSATS, which does not contain the flexible ACP domain or the linker region between the AT and ACP domains.

4.3.1 Crystallisation of module 5's KS-AT didomain

M5_KSATS was overproduced and purified as described in section 7.3. Screening in 96-well sitting drop crystal trays, containing JCSG Core I and II screens, produced a number of crystals (Figure 4.4). Cryogenic freezing and diffraction on beamline I06 at Diamond

Figure 4.4. *Crystals obtained for M5_KSATS.*

Light Source showed the crystals obtained from conditions E9 and H2 of Core I contained protein (Figure 4.5). Unfortunately, the small crystals only diffracted to 6 - 8 Å, meaning further optimisation would be required for the crystals to be of sufficient quality to obtain a high-resolution structure.

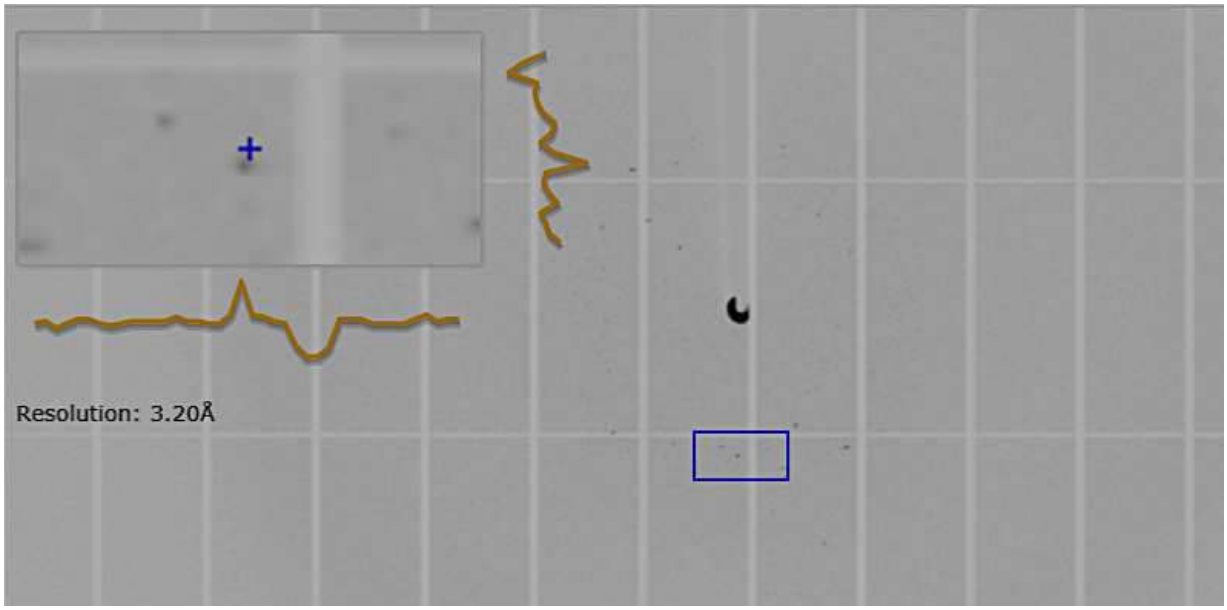


Figure 4.5. *Diffraction pattern of a M5_KSATS crystal produced under condition H2 of Qiagen Core I.*

Further optimisation, based upon the conditions containing protein crystals was then planned. The trials were based extensively around 4-(2-hydroxyethyl)-1-piperazineethane-sulfonic acid (HEPES) and sodium citrate buffers, with Polyethylene glycol (pEG) 3350, due to their presence in a number of promising conditions. Crystal trials were set up in sitting drop wells with 1:1, 2:1 and 3:1 (sample to buffer) ratios. Wells contained either 0.1 M HEPES or 0.1 M sodium citrate, both at pH 6.5, with 20% pEG 3350 and a range of

different salts (NaCl, $CaCl_2$, HK_2O_4P, $MgSO_4$, ClK, H_2KO_4P and HNa_2O_4P, all 0.2 M).

Further optimisation of salt levels (0.05 - 0.5 M) and pEG 3350 concentrations (17 - 23 %) were carried out around the most promising conditions at both room temperature and 4°C, at a range of protein to buffer concentrations (1:2, 1:1, 2:1 and 3:1) and in sitting and hanging drop trays. pH was also adjusted for some conditions (pH 6.0 - 8.5) and sample concentration was varied (2 - 10 mg/mL). However, no significant improvements upon the original crystals were seen, potentially due to unfavourable entropy.

Since M5_KSATS was unexpectedly monomeric, crystallisation trials were repeated using full-length M5_KSAT, which is dimeric and contains the KS-AT didomain and the AT-ACP linker region. Crystallisation trials were set up using all four cores of JCSG Core suite in 96-well sitting drop plates, but no improvements on the crystals produced for M5_KSATS were found.

4.4 Probing the interactions between M5_KSAT and M5_ACP using carbene footprinting

All work in section 4.4 was done by Dr. Matthew Jenner using the previously produced homology models. Carbene footprinting was used in an attempt to map interactions onto homology models of M5_KSAT and M5_ACP as previous work had failed to provide structural information on the domains of M5.

Carbene footprinting is a recently developed technique that provides information on which residues are involved in protein-protein interactions.[183] Protein labelling is achieved via incubation with an aryl diazirine reagent, which upon laser irradiation at 350 nm liberates the N_2 and yields a highly reactive carbene species *in situ*. The carbene has the ability to insert into C-H bonds of non-polar amino acid sidechains, such as in trypto-

phan, and the O-H, N-H or S-H of polar amino acid sidechains, such as in lysine and arginine, thereby providing a wide labelling coverage of the solvent-exposed surface. Protein samples are then enzymatically digested, using trypsin, and the resulting peptides are examined by mass spectrometry. Comparison of the isolated domain with labelled complexes will then allow the deciphering of interaction surfaces from the disappearance of visible peptides (Figure 4.6).

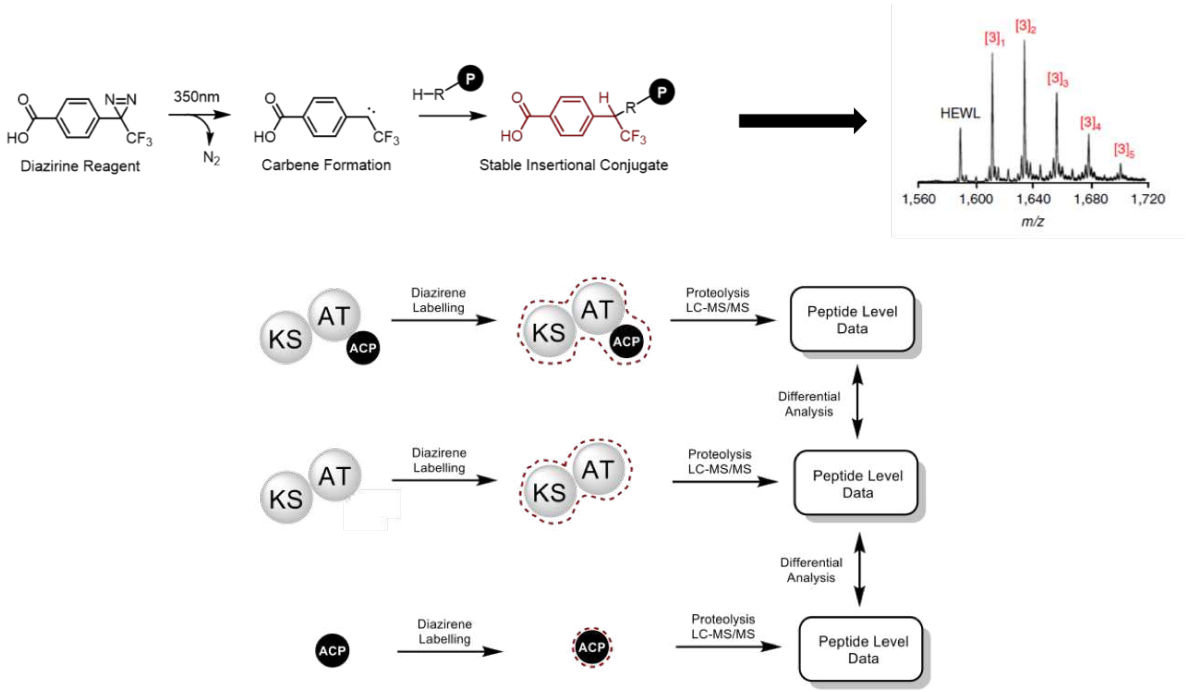


Figure 4.6. *The workflow process of carbene footprinting for hen egg white lysozyme (HEWL, top).*[183] *Carbene footprinting allows the identification of residues masked as a result of protein-protein interactions (bottom).*

M5, M5_KSAT and M5_ACP were diazirene labelled and enzymatically digested. No labelling was observed on the KS or AT domains for both M5_KSAT as a stand-alone construct, or as part of the complete module. This indicates that the corresponding residues were inaccessible, perhaps due to aggregation. In contrast, labelling was observed on the three peptide fragments of the ACP domain for both the standalone M5_ACP construct and the entire M5, suggesting that the ACP domain was not aggregated.

Residues 114-121 of the ACP domain exhibited similar levels of labelling in the presence and absence of the rest of M5 (Figure 4.7), suggesting that they were not involved in any intramodular protein-protein interactions. Due to its location at the C-terminus of the ACP's final helix, it perhaps plays more of a role in interactions with the downstream post-PKS enzymes. Diazirine labelling of the peptides corresponding to residues 53 - 66 and 101 - 113 of the ACP domain was significantly masked in M5, compared to M5_ACP, suggesting these regions were involved in intramodular protein-protein interactions or

were involved in the module's aggregation.

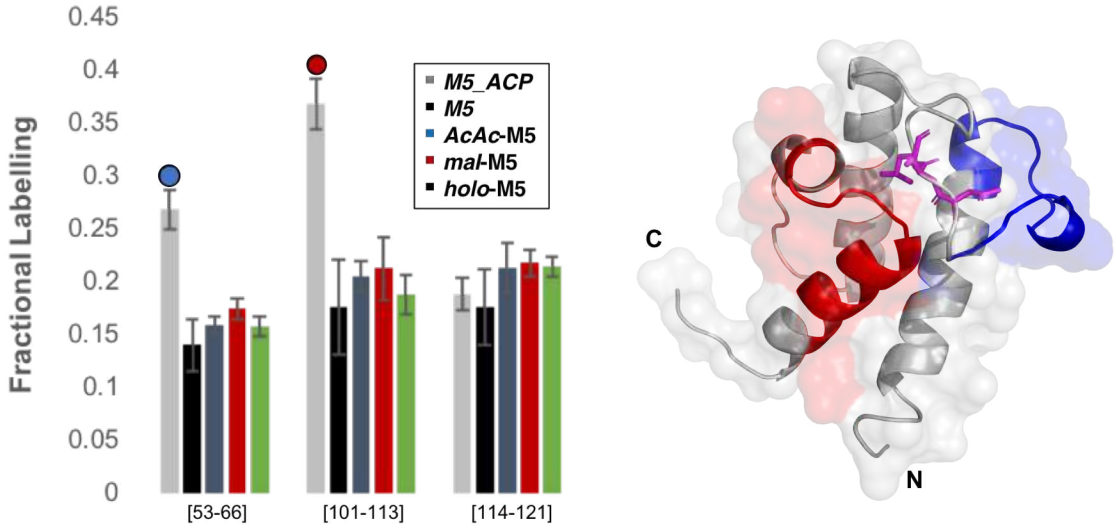


Figure 4.7. *The fractional labelling seen for M5_ACP peptides after AspN digests (left). A homology model (Phyre) highlighting the two peptides (53 - 66 and 101 - 113) significantly masked in M5 (right), along with the substrate attachment site (DSL).*

M5 was also diazirene-labelled in its *holo*, malonyl-*holo* and acetyl-acetyl-*holo* forms, all resulting in no labelling of the KS-AT regions and no statistically significant deviation from the labelling of *apo*-M5.

4.5 Structural motifs involved in the dimerisation of module 5

As mentioned in section 2.3.3, although M5 and M5_KSAT are dimeric, M5_KSATS (without the AT-ACP linker region) is monomeric. This is surprising as module dimerisation is virtually reliant on interactions between KS domains, which have large interaction interfaces. Although in this system, KS domain dimerisation is certain to play a role in module dimerisation, it is probable that a second dimerisation motif exists in the linker region.

There are other systems in which the KS domain is not solely responsible for modular dimerisation. DH domains, for example, also usually purify as dimers and can play roles

in modular dimerisation.[123] Additionally, it was also recently suggested that the ACP domain in PikAIII contains a second dimerisation motif since the removal of the ACP domain results in a loss of dimerisation.[117] Although ACP docking domains have previously been shown to dimerise[124] and could therefore play this role, M5_ACP has no such docking domain as there is no need to facilitate interactions with a downstream KS.

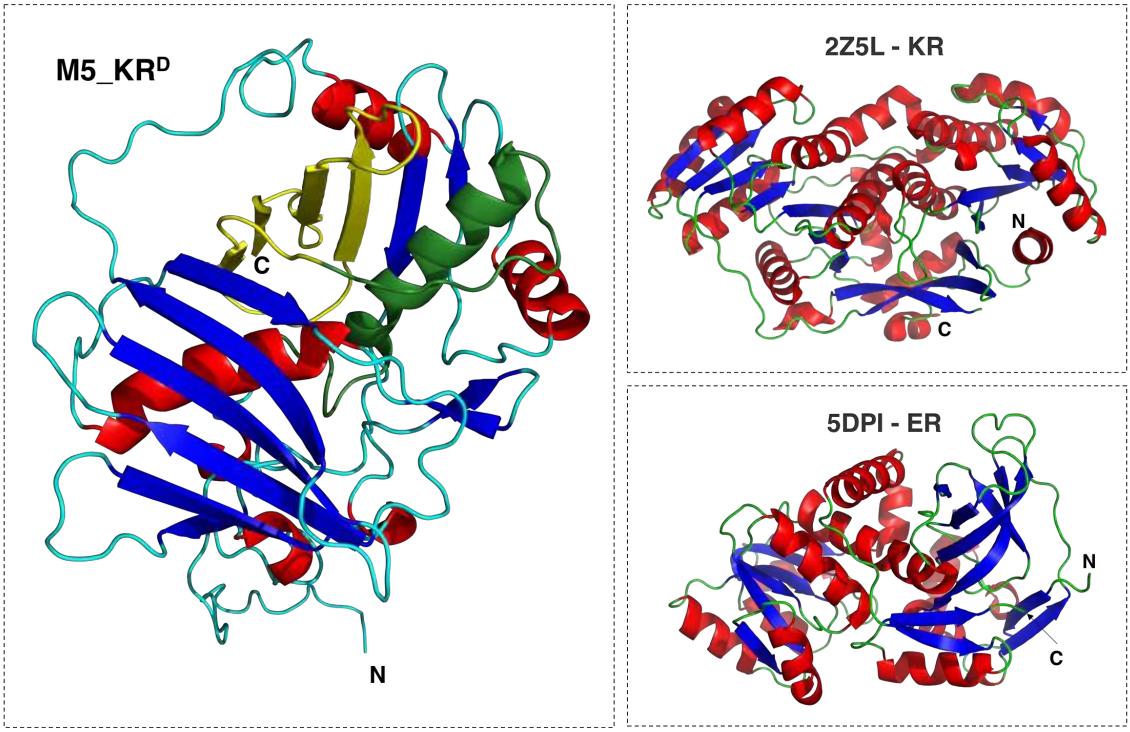


Figure 4.8. *A homology model of M5's AT-ACP linker region (left) and crystal structures of a KR (top right) and ER (bottom right) domain. The highlighted α-helix of M5-KR^D is essential for module dimerisation and the β-strands are not.*

A homology model was created for the AT-ACP linker region to search for the presence of known structural motifs (Figure 4.8). One identified motif was a partial Rossmann fold that formed an extended β-sheet, suggesting the linker region may contain remnants of a historical KR or ER domain. One of the most common super-secondary structures, Rossmann folds consist of a nucleotide-binding $\beta\alpha\beta$ fold, with the possible addition of extra β-strands to create an extended β-sheet. Rossmann folds show a low level of sequence similarity between families, with the only widely conserved motif being a GxGxxG/A motif located in the tight bend between the first β-strand and α-helix. Dimerisation has been noted between Rossmann folds in a number of solved structures, such as dehydrogenases (PDB: 1a71, 1A4I),[184;185] indicating the regions could play a role in module dimerisation.

In order to investigate the boundaries of the dimerisation interface, three C-terminal (Fig-

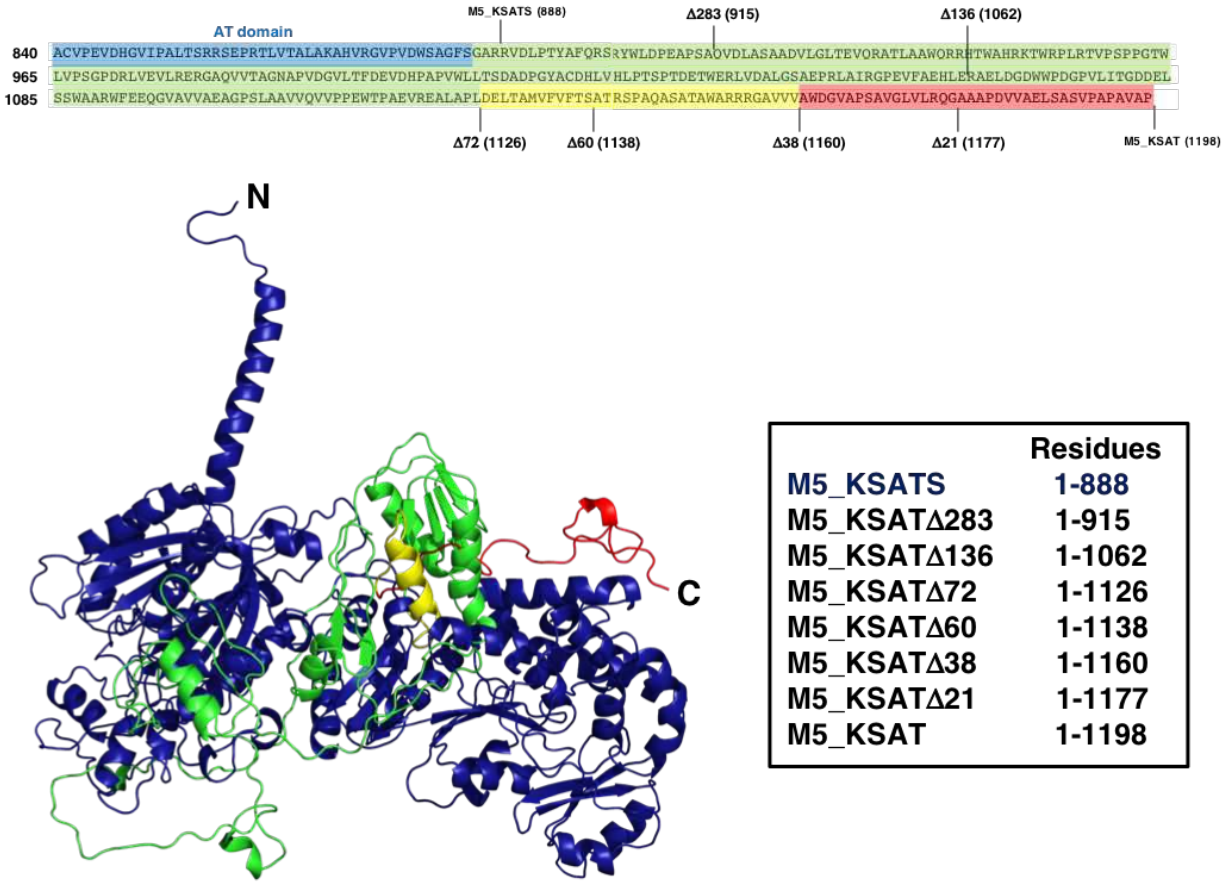


Figure 4.9. *Sequence (top) and a homology model (bottom left) showing the termination points of various truncated M5 constructs (bottom right).*

ure 4.9) and four N-terminal truncated constructs were overproduced. Of the three C-terminal truncated constructs, M5_KSATΔ283 (residues 1 - 915, 100,970 Da) and M5_KSAT Δ136 (residues 1 - 1062, 117,115 Da) almost completely aggregated and M5_KSATΔ72 (residues 1 - 1126, 123,947 Da) corresponded to a monomeric eluted protein (Figure 4.10). As this construct omits 72 residues from the C-terminal of M5_KSAT, this suggests that these residues were essential for module dimerisation. Three further intermediate constructs were overproduced, with M5_KSATΔ38 (residues 1 - 1160, 127,548 Da) and M5_KSATΔ21 (residues 1 - 1177, 129,231 Da) being soluble and M5_KSATΔ60 insoluble (residues 1 - 1138, 125,285 Da). M5_KSATΔ38 and M5_KSATΔ21 both corresponded to dimeric protein complexes, suggesting the loss of dimerisation occurs between residues 1126 and 1160. As this region only contains an α-helix, it can be assumed that the α-helix plays a vital role in module dimerisation. It can also be concluded that the final three β-strands that are proposed to form part of an extended β-sheet are not essential to module dimerisation. The α-helix in question is located within a hydrophobic pocket, meaning the large amount of aggregation seen for M5_KSATΔ38 could be due to the removal of the β-sheet exposing the helix, resulting in the protein unfolding.

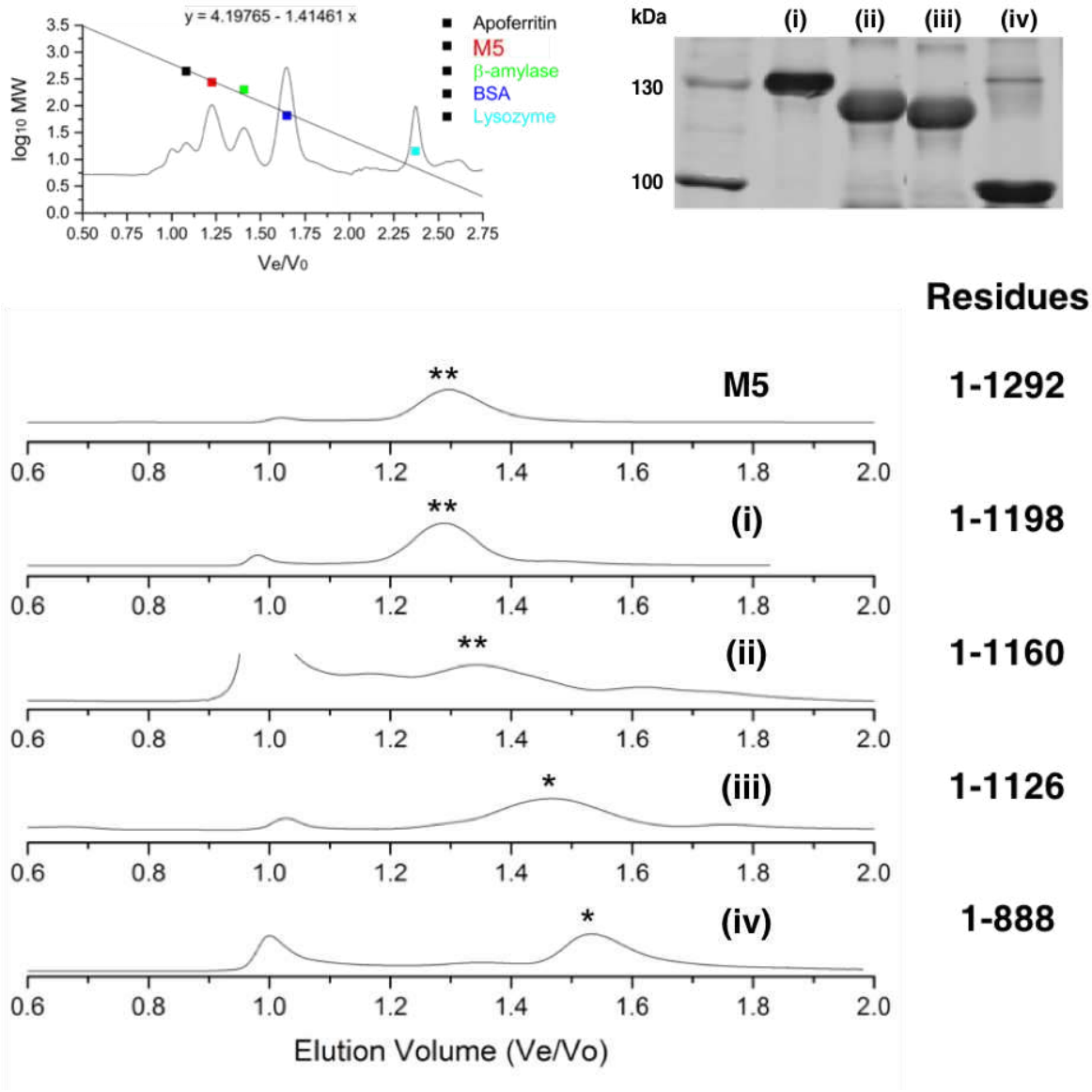


Figure 4.10. *Gel filtration chromatograms of Superdex 200 calibration (top left), SDS-PAGE analysis of overproduced truncated M5 constructs (top right) and corresponding gel filtration (bottom) that formed either monomers (*) or dimers (**) of M5 (141 kDa), M5_KSAT (**(i)**, 131 kDa), M5_KSATΔ38 (**(ii)**, 128 kDa), M5_KSATΔ72 (**(iii)**, 124 kDa) and M5_KSATS (**(iv)**, 98 kDa).*

Sequence analysis indicated that the linker region contained significant structural similarity to known KRs (e.g. from spinosyn module 3), with several domain finding algorithms labelling the linker region as a KR domain. Although M5's linker region is too short to be a complete KR, and no reduction of the β-keto group occurs, the region contains a conserved tryptophan, indicative of an A-type KR, and an active-site tyrosine. The proposed role in module dimerisation has led to the region being termed a KR dimerisation motif (KR^D).

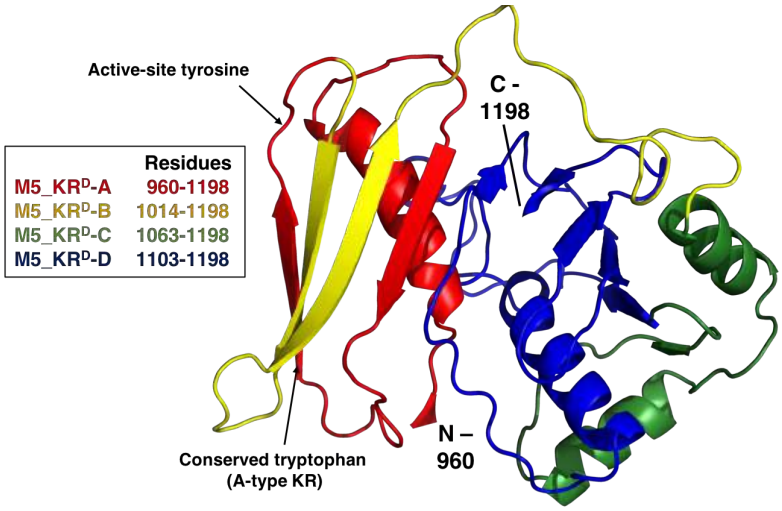


Figure 4.11. *A homology model showing the predicted structure of overproduced constructs containing various lengths of M5's AT-ACP linker region.*

N-terminal truncations (Figure 4.11) showed M5_KR^D-A (residues 960 - 1198, 29,036 Da), M5_KR^D-B (residues 1014 - 1198, 23,319 Da) and M5_KR^D-C (residues 1063 - 1198, 17,825 Da) to be homodimers (54, 42 and 34 kDa, respectively), whereas M5_KR^D-D (residues 1103 - 1198, 13,526 Da) completely aggregated (Figure 4.12). Each dimerising construct contained at least two α-helices, one of which was previously shown to be essential for module dimerisation, and a β-sheet that could potentially provide an interaction interface with the corresponding β-sheet of a dimeric partners' KR^D.

X-ray crystallography of the AT-ACP linker region was then attempted. M5_KR^D-A was chosen as the ideal construct for crystallisation due to a low level of aggregation. To increase the chances of crystallisation, the His-tag was cleaved with TEV (tobacco etch virus)-Sfp (surfactin phosphopantetheinyl transferase) before removal of the TEV-Sfp and uncleaved proteins by reverse NTA column. The cleaved protein was then further purified by gel filtration and the fractions containing the protein were transferred into a buffer of 20 mM Tris, 50 mM NaCl, pH 7.4 by PD10 column. The protein was concentrated to 10 mg/mL and crystallisation conditions were tested using Qiagen cores I and II. Unfortunately, no crystals were obtained.

Spirotetronate producing PKS pathways ubiquitously end with standalone modules responsible for a non-reductive elongation. Sequence alignments with the final modules of other spirotetronate PKSs were carried out (Figure 8.3), with a number of these modules shown to contain significantly shorter AT-ACP linker regions ($<$ 100 residues) and no

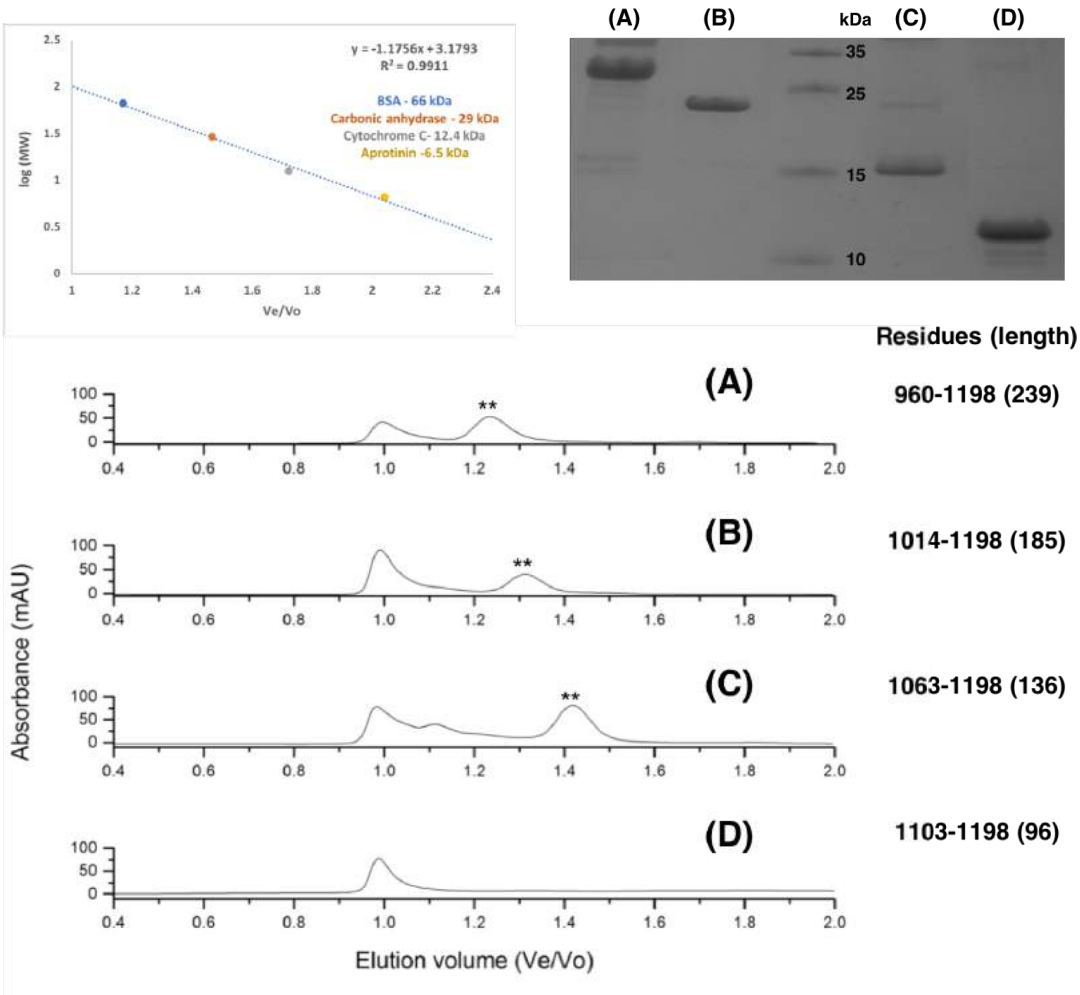


Figure 4.12. *Calibration of the Superdex 75 (top left), SDS-PAGE analysis of overproduced KR^D proteins (top right) and gel filtration chromatograms of constructs containing varying lengths of M5's AT-ACP linker region (bottom) that formed dimers (**) or completely aggregated ($V_e/V_0 = 1$). KR^D-A (**A**, 29 kDa), KR^D-B (**B**, 23 kDa), KR^D-C (**C**, 17 kDa) and KR^D-D (**D**, 14 kDa).*

remnants of Rossmann folds. However, the final modules of the chlorothricin and kijanimicin PKS pathways (ChlA6 and KijS5) contain AT-ACP linkers that display significant sequence similarity/identity to M5 (71/60 and 59/50 %, respectively). Although the linkers for ChlA6 and KijS5 are substantially longer ($\sim$ 510 and 390 residues, Figure 4.13), they also contain regions predicted to be Rossmann folds.

KijS5 has previously been reported to contain a non-elongating KR and ChlA6 to be a minimal module.[147;186] Although the regions were too short to contain an entire KR, several domain picking algorithms predicted the linker regions as KR domains. ChlA6 also contained a conserved LDD motif, indicative of a B-type KR, and a GGxGxxG NADP-binding motif (GGPGVTG),[187] further suggested it was a KR remnant. Further sequence analysis showed that ChlA6 and KijS5's AT-ACP linker regions contained a dimerisation element

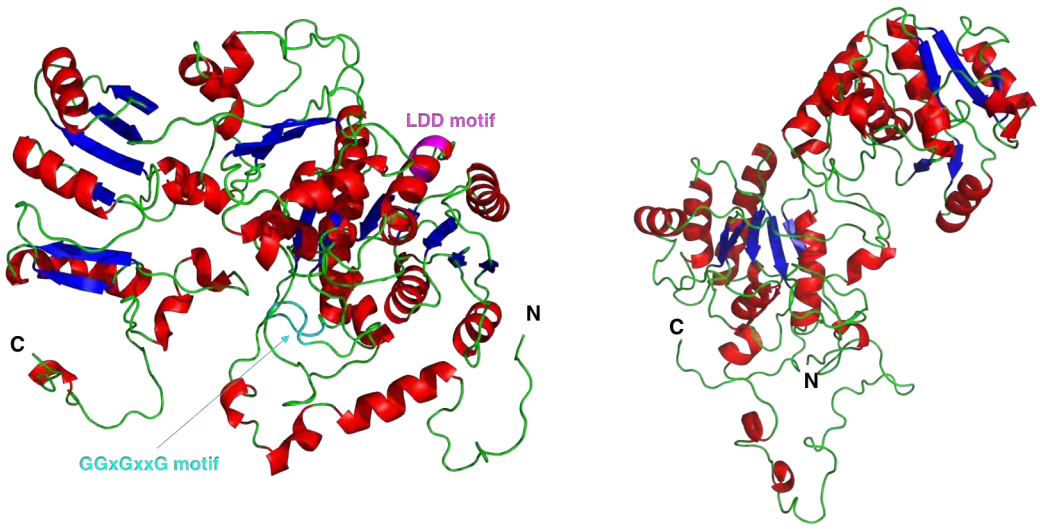


Figure 4.13. *Homology models of the AT-ACP linker region of ChlA6 (left) and KijS6 (right). α-helices and β-strands are displayed. Models were based upon the following structures: 2vz8, 6fik, 6c9u (ChlA6 and KijS5)*

found between the AT and KR domains of various PKS systems,[122] followed by the remnants of a KR, leading to the regions also being labelled as KR^Ds. The same dimerisation element was not present in M5, suggesting the presence of a novel dimerisation motif within M5-KR^D.

4.6 The morphology of module 5

Cryo-EM was first successfully applied to the study of PKSs in 2014, when cryo-EM structures were obtained for the penultimate module of pikromycin PKS (PikAIII, β-module) in each conformation of its catalytic cycle.[123;188]

4.6.1 Sample optimisation of module 5

The folding of M5 under various conditions was first confirmed using circular dichroism (CD), before further examination at varying temperatures showed a significant reduction in absorbance at 208 nm between approximately 35^oC and 50^oC. This indicates that M5 is stable at room temperature over a short time period (2 hours) and starts to unfold at approximately 35^oC (Figure 4.14).

Negative-stain EM was used to obtain preliminary images to assess the quality of the

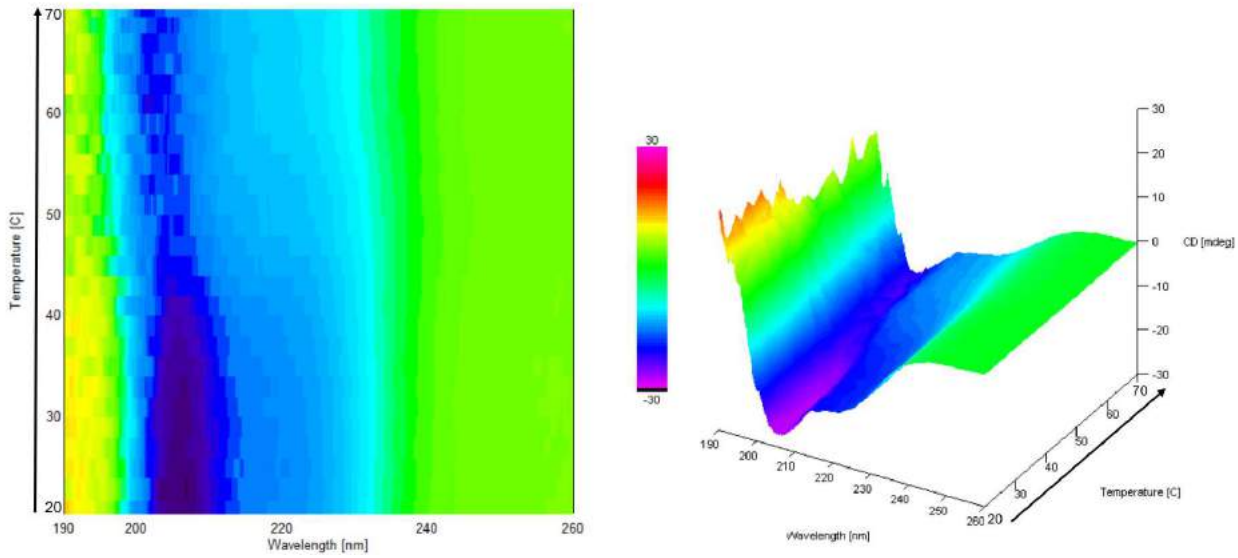


Figure 4.14. *Variable-temperature CD spectra of M5 at pH 6.5.*

sample. The advantage of negative-stain EM is that the background is stained with a heavy metal, creating a contrast between the sample and the background without staining of the sample. This ensures that the conformation of the protein is not affected by the staining procedure, whilst providing the required contrast for images of the sample to be collected. Although negative-stain is suitable for gaining a base level understanding of the sample, cryo-EM would be needed to reach the levels required to gain any significant structural information.

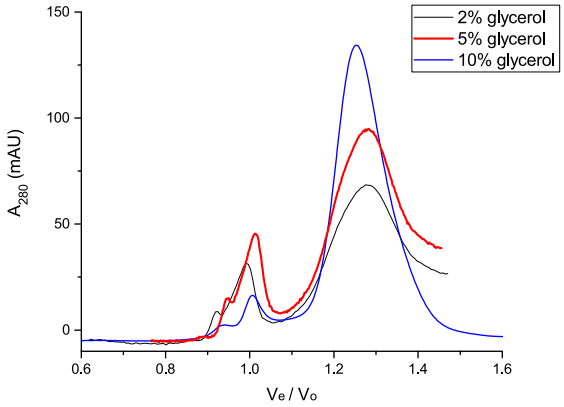


Figure 4.15. *Gel filtration chromatograms for M5 (S200, 20 mM Tris, 150 mM NaCl and varying levels of glycerol).*

Gel filtration was used to assess the level of aggregation of M5 under various buffer conditions, with optimal conditions found to be 20 mM Tris, 150 mM NaCl, 10 % glycerol, pH 7.4. To acquire cryo-EM pictograms, it is advantageous to have lower glycerol levels to prevent a loss of contrast. Therefore, further optimisation was carried out in which lower glycerol levels were tested (2 % and 5 %). Both resulted in significant aggregation (Figure 4.15). Therefore, 10 % glycerol was used in all buffers.

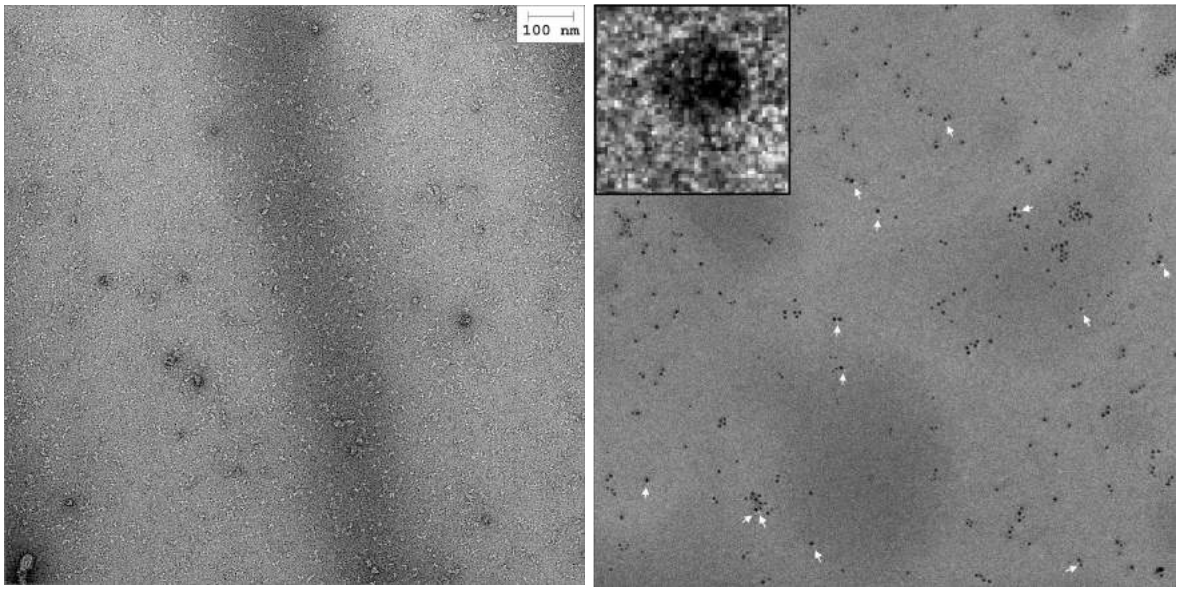


Figure 4.16. *Negative-stain electron microscopy image of M5 (left, light areas). A cryo-EM visualisation of M5 with Ni-NTA-Au (right, Daniel Griffiths). Each black dot corresponds to an M5 homodimer with Au bound to each of the KS docking domains.*

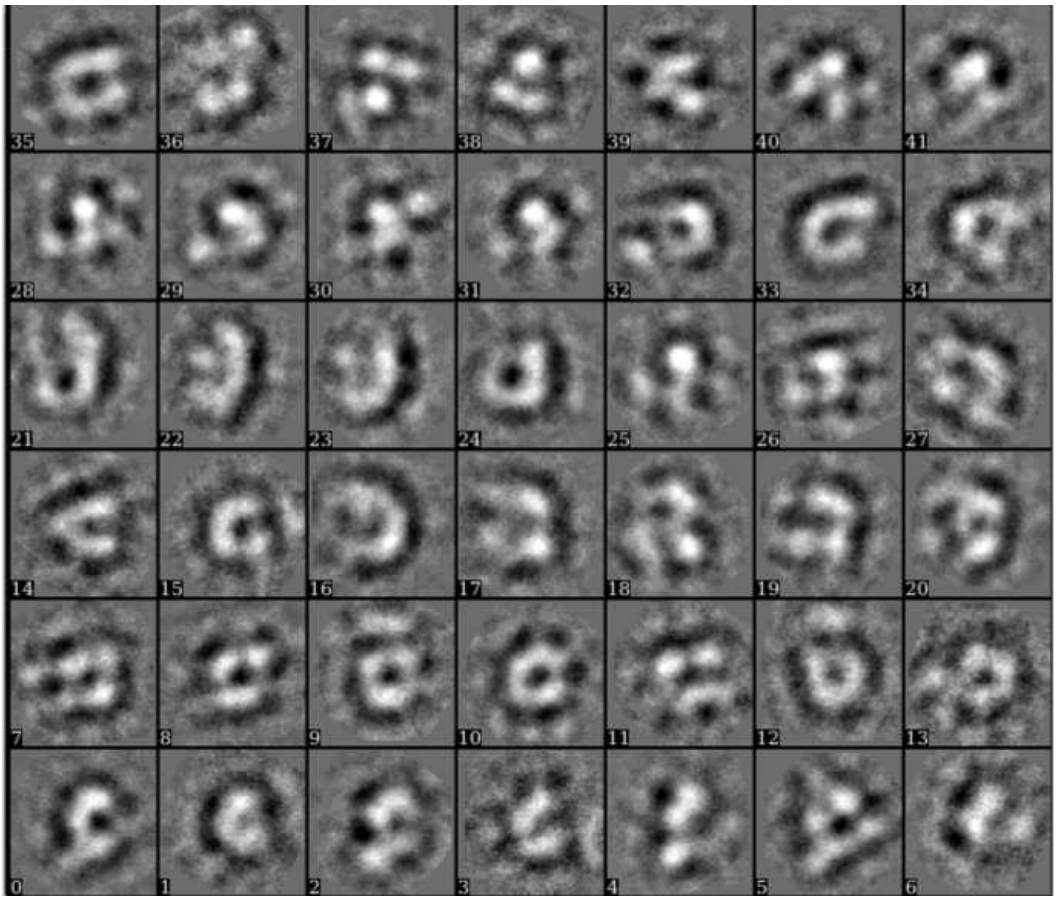


Figure 4.17. *2-D cryo-EM reconstructions of individual M5 dimers.*

Negative-stain EM highlighted M5's propensity to aggregate (Figure 4.16 (left)). As a re-

sult, other work was prioritised and acquisition of a morphological map for M5 was followed up by Dr. Daniel Griffiths, a postdoctoral researcher in the group, based at Monash University. Cryo-EM images were obtained of unconcentrated fresh protein samples, directly off the gel filtration column, in an attempt to limit aggregation. These images showed aggregation had decreased significantly (Figure 4.16 (right)) and 2-D particle averaging indicated the observed particles were of approximately the same size in different orientations (Figure 4.17), highlighting the homogeneity of the sample. However, the images acquired were not of high enough quality to consider producing a 3-D morphology model of M5.

Examination of the 2-D reconstructions led to some interesting observations. A significant number of the reconstructions showed M5 to form a central chamber, similar to what was noted for PikAIII. One end of the observed chamber is wider in each case, likely corresponding to two KS domains with a large interaction interface (Figure 4.18). Either side of the reaction chamber is likely to be the AT domains, which are proposed not to associate in PKS systems, followed by the KR^D region with a contact region corresponding to the dimerisation motif (section 4.5). The ACP is likely to be omitted from these reconstructions due to its structural mobility. The other reconstructions are likely to correspond to various orientations of the same structural conformation.

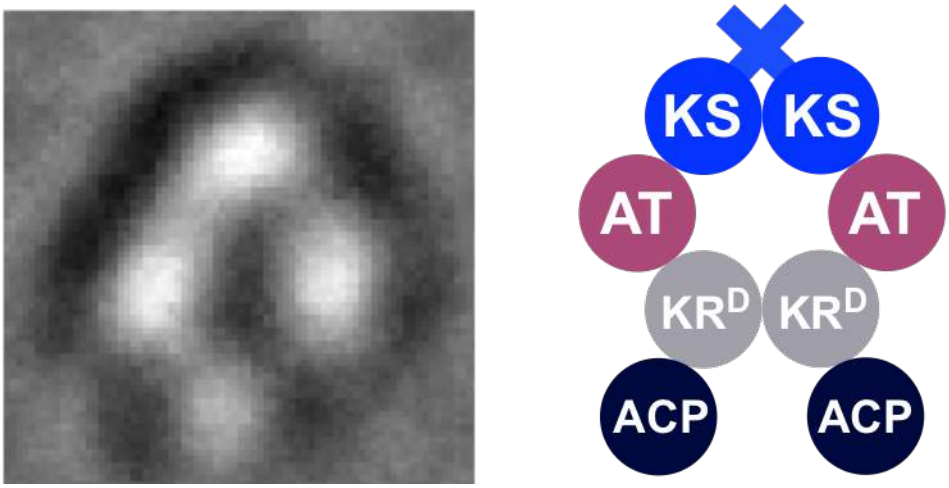


Figure 4.18. *A 2-D cryo-EM reconstruction of a M5 dimer (left) and a schematic representation (right).*

4.7 Conclusion

Structural elucidation of M5's components was attempted through a variety of approaches, including NMR, X-ray crystallography, cryo-EM, carbene footprinting and gel filtration. Most of these approaches were hindered due to the high propensity of M5 to aggregate. Sample optimisation of M5_ACP resulted in suitable samples for further structural elucidation and sample optimisation of the KS-AT didomain prevented aggregation and allowed the obtaining of small crystals. Although diffraction was observed for two crystals, they did not provide the necessary resolution to acquire a structure (6 - 8 Å).

Carbene footprinting of M5, using homology models, identified two regions on the ACP that interact with the rest of the module. However, no labelling was observed in the KS and AT domains, suggesting aggregation had occurred. This is possibly indicative of the formation of a higher order structure involving the KS and AT domains, leaving the ACP free to interact with the post-PKS enzymes.

Analytical gel filtration showed that the removal of a linker region between the AT and ACP domains resulted in a loss of dimerisation. Further intermediate constructs showed dimerisation was lost between residues 1126 and 1160, a region containing a long α-helix, suggesting this region plays a vital role in module dimerisation. N-terminal truncated constructs were also produced to determine the start of the region essential for module dimerisation. Analytical gel filtration of three different constructs showed they were dimers. Importantly, each of these three constructs contained the α-helix shown to be essential for module dimerisation, as well as a β-sheet. It was subsequently hypothesised that the linker regions long β-sheet was able to pair to a dimerisation partner to provide an interaction interface, supporting a dimerisation element formed by the intermodular linking of α-helices.

Sequence analysis of the linker region suggested it contained the remnants of a KR domain. AT-ACP linker regions were also noted in ChlA6 and KijS5, both of which contained the conserved residues of a previously noted dimerisation element not present in M5-KR^D.[122]

Finally, cryo-EM images showed single particles of M5 were observed from various orientations. A chamber was observed in a significant number of the reconstructions, with what appears to be a large KS-KS dimerisation interface on one side and a smaller contact region on the other side. This was proposed to be the hypothesised dimerisation motif located towards the end of the AT-ACP linker region.

4.8 Future Work

Structural elucidation of the AT-ACP linker region would provide information on the dimerisation motif. Both X-ray crystallography and NMR would be potentially viable ways to understand the linker region functions. The homologous region of ChlA6 could also be targeted, allowing comparisons with M5's linker.

Using cryo-EM to elucidate the structure of M5 in its varying conformations will be the second priority. Further optimisation may provide a suitable sample to allow this. Acquisition of approximately two hundred 2-D class averages representing a range of particle orientations would allow for the creation of a de novo 3-D model, before refinement of the resulting structure could provide a morphology map of M5.

Successful acquisition of a morphology model would provide motivation to resume the acquisition of a high-resolution structure of M5_ACP. 3-D spectra (cbcaconh, cbcanh, ...) would first be acquired, allowing complete backbone assignment. Distance restraints could then be acquired through a ^{15}N Nuclear Overhauser Effect Spectroscopy (NOESY) HSQC from which an experimental 3-D structure of M5_ACP could be acquired and mapped into M5. Any solved structure of the AT-ACP linker region and a homologous KS-AT didomain could then be mapped into M5, creating a pseudo-atomic level structure of the entire module in each step of its catalytic cycle.

4.9 A homologous gene cluster in *Amycolatopsis albispora*

A blast search of the QMN biosynthetic gene cluster highlighted close similarities with a region of *Amycolatopsis albispora*'s gene cluster (Figure 4.19). Homologues of all genes known to play a role in the production of QMN were present (Figure 4.20, 83 - 96 % identity), with *qmn*M and *qmn*N the only genes without identified homologues. Homologous genes to those encoding the QMN PKS pathway were present (88 - 93 % identity), with all docking domains completely conserved. Genes predicted to encode a number of NRPS domains were present. Modification of the QMN precursors by the NRPS domains could result in the production of modified precursors with the potential to form QMN skeletons with attached functional groups.

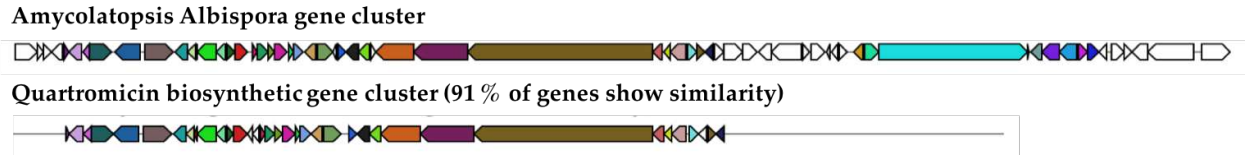


Figure 4.19. *A comparison of a gene cluster from Amycolatopsis albispora (top) and the homologous quartromicin biosynthetic gene cluster (bottom).*

As the structural elucidation of M5 was problematic, the homologous protein from *Amycolatopsis albispora* could be studied instead. Homology models predicted the protein to be structurally very similar to M5 (91 / 86 % similarity/identity) and to contain a similar AT-ACP linker region (Figure 4.21). Sequence alignments of both AT-ACP linkers (77 / 68 %) highlighted the strong similarity and led to the region of *Amycolatopsis albispora* being categorised as a KR^D.

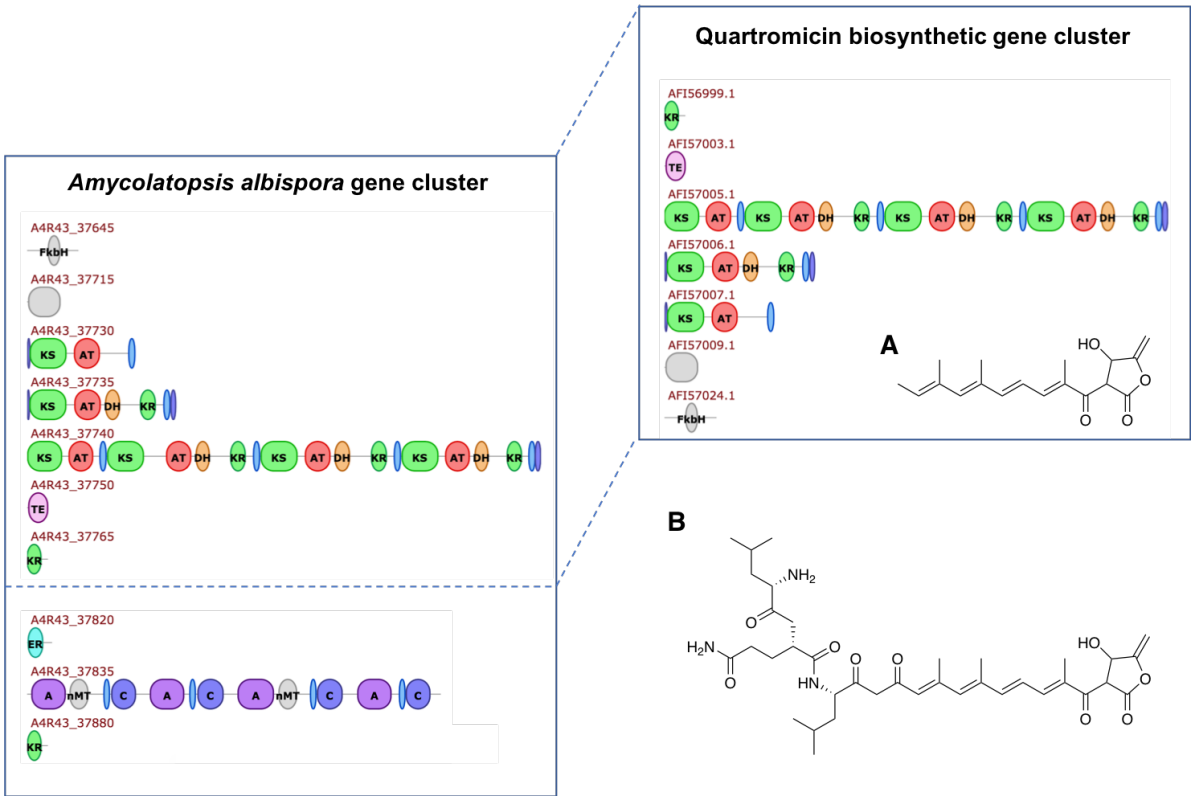


Figure 4.20. *Proteins hypothetically encoded by the Amycolatopsis albispora (left) and the QMN gene cluster (right). The resulting predicted precursors for quartromicin (**A**) and the Amycolatopsisnalbispora gene cluster (**B**) are also shown.*

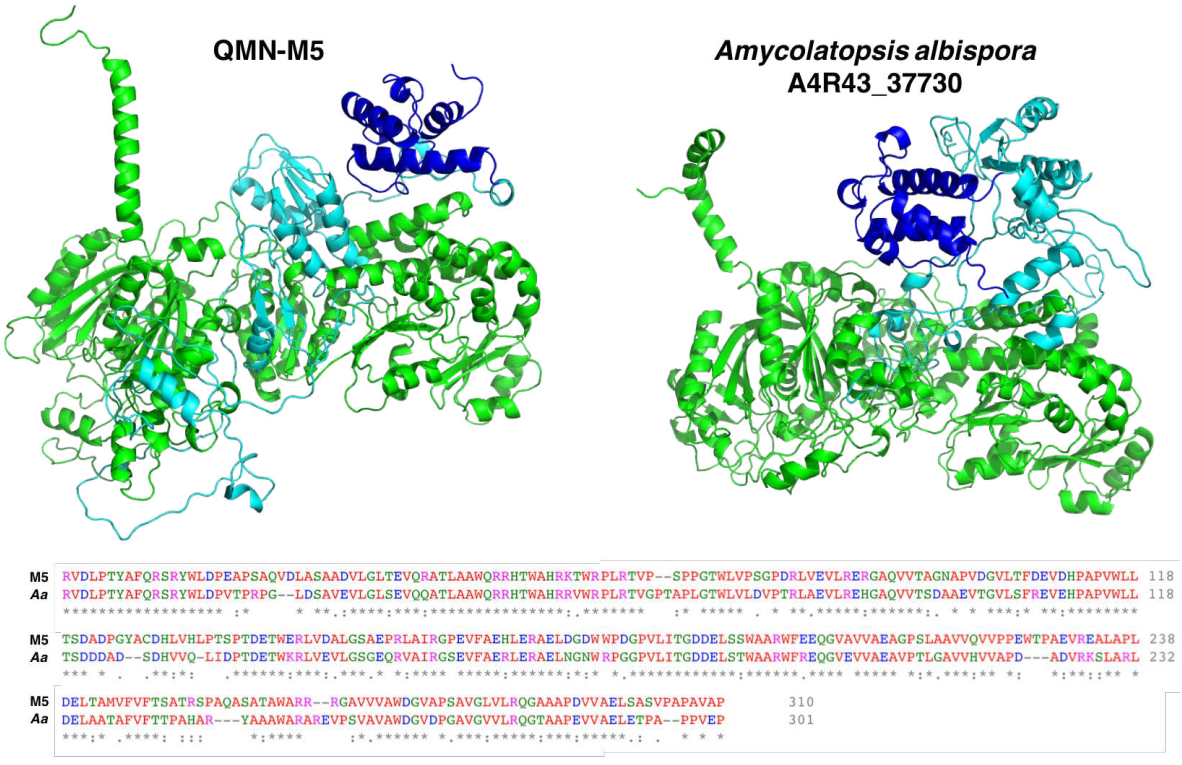


Figure 4.21. *Homology models of QMN module 5 (top left), the homologous module from Amycolatopsis albispora (top right) and a sequence alignment between the two modules (bottom). Homology models were produced using 2vz8, 6gik and 6c9u in both cases.*

Chapter 5

The intermodular interactions of quartromicin module 5

5.1 Acyl transfer between upstream ACPs and quartromicin module 5

An important aspect of the QMN PKS pathway is the intermodular interactions between M3/M4 and M5. The ability of M5 to successfully accept, extend and modify nascent polyketides from two different ACPs provides an interesting opportunity to study the interactions between the domains. Understanding the promiscuous selectivity of M5's KS, as well as M3's ACP, may help understand how pathway fidelity is ensured.

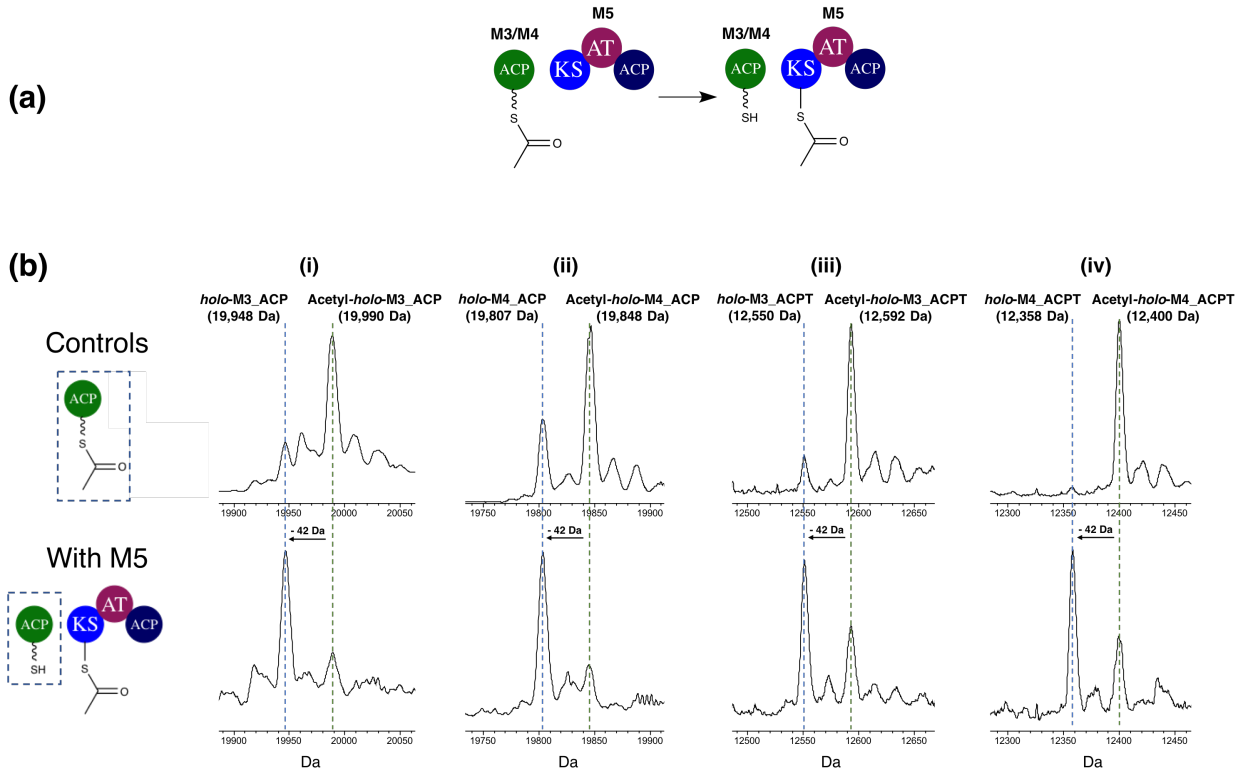


Figure 5.1. *Schematic representation of acyl transfer from an upstream ACP to a downstream module **(a)**. Deconvoluted LC/MS chromatograms **(b)** monitoring the protein in the box (left). ACPs used were **(i)** M3_ACP, **(ii)** M3_ACP, **(iii)** M3_ACPT and **(iv)** M4_ACPT.*

Since M3_ACP and M4_ACP appear to contain a C-terminal docking domain, we first set out to determine whether these docking domains are essential for acyl transfer from the ACP domains to the KS of module 5. The full-length (M3_ACP and M4_ACP) and truncated (M3_ACPT and M4_ACPT) upstream ACPs were loaded with acetyl-CoA, as previously described (chapter 3), before being incubated overnight with buffer (control) or equimolar quantities of M5. Intact protein mass spectrometry was then used to monitor

the loss of an acetyl group from the ACPs. A small amount of *holo*-ACP can be detected in all the control experiments due to hydrolysis of the acyl chain. A significant difference was noted in the presence of M5. Almost complete acyl transfer was observed for each full-length ACP (Figure 5.1), indicating that both ACPs could transfer a substrate to M5. Truncation of the docking domain resulted in a decrease to approximately 80 % substrate transfer, suggesting that whilst the docking domains may play a role in intermodular interactions, they are not essential for substrate transfer.

5.1.1 Calculating the rate of acyl transfer

Assays were then repeated with aliquots taken over a series of time points to better understand the impact of the docking domain on the rate of transfer (Figure 5.2). The rate of transfer was noticeably quicker for M3_ACP than M4_ACP. Interestingly, whilst the removal of the docking domain had a significant effect on the rate of transfer from M3_ACP, the rate of transfer from M4_ACP was unaffected. In fact, transfer from M4_ACP and the ACP constructs only containing the four conserved helices (M3_ACPT, and M4_ACPT) all proceeded at a similar rate. This suggests M3's ACP docking domain (M3_^{ACP}dd) more efficiently facilitates acyl transfer than M4's (M4_^{ACP}dd). For example, after 3 hours of monitoring the reaction, approximately 15 % transfer was seen from M3_ACP and 10 % transfer was noted from M3_ACPT, M4_ACP and M4_ACPT.

5.1.2 The stoichiometry of intermodular acyl transfer

Assays were repeated over 90 minutes with varying concentrations of M5 to ACP. Controls were run with equimolar quantities of M5-C207A. Increased levels of transfer off each ACP was observed as the concentration of M5 increased and more acyl transfer was observed from M3_ACP than M4_ACP for each concentration (Figure 5.3). Significantly, the amount of transfer for 1:1 ratios of each ACP to M5 were as expected given previous results (section 5.1.1).

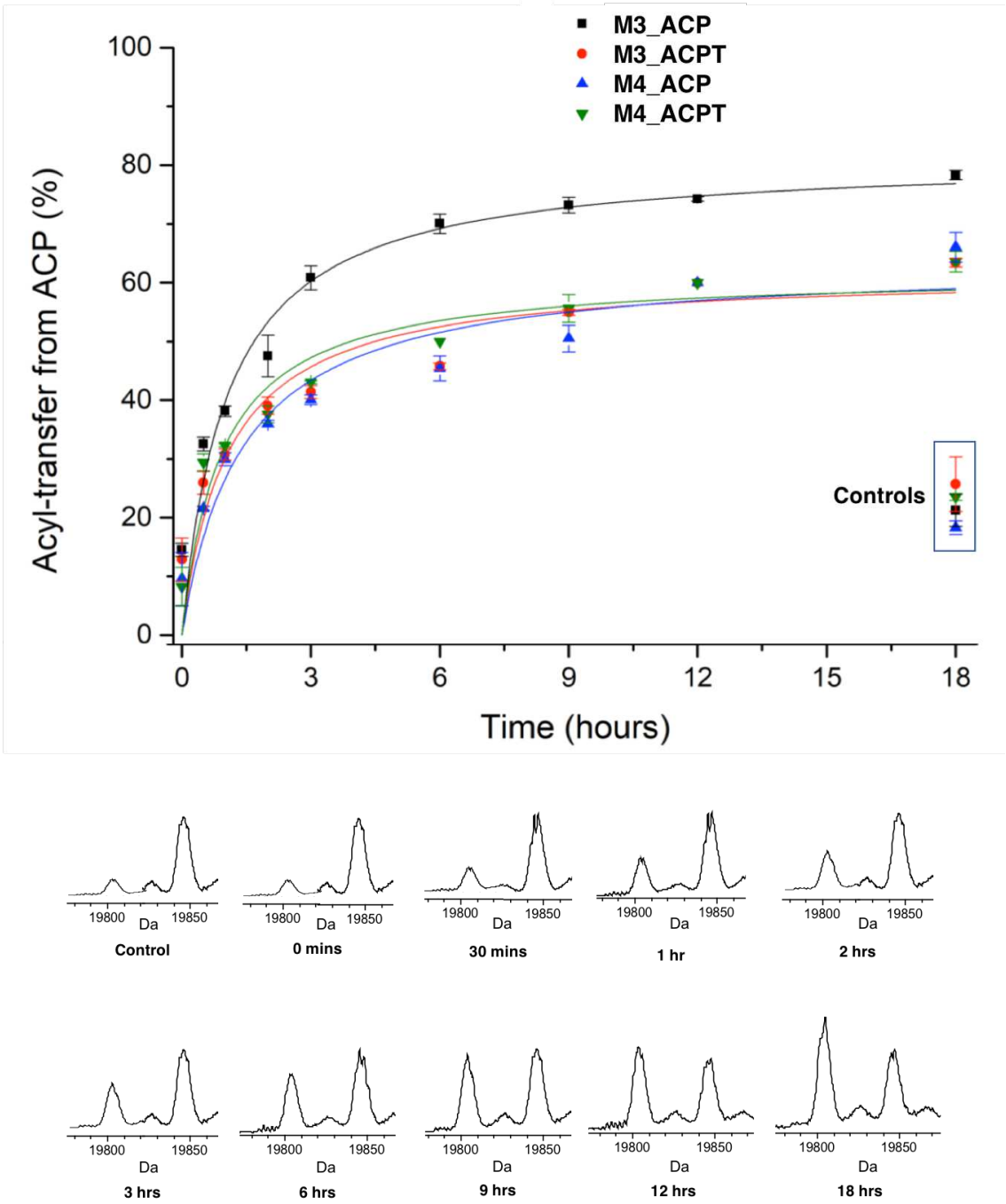


Figure 5.2. *A graphical representation of acyl transfer to M5 over time (top) and corresponding deconvoluted spectra for M4_ACP (bottom). Experiments were run in triplicate and errors correspond to one standard deviation.*

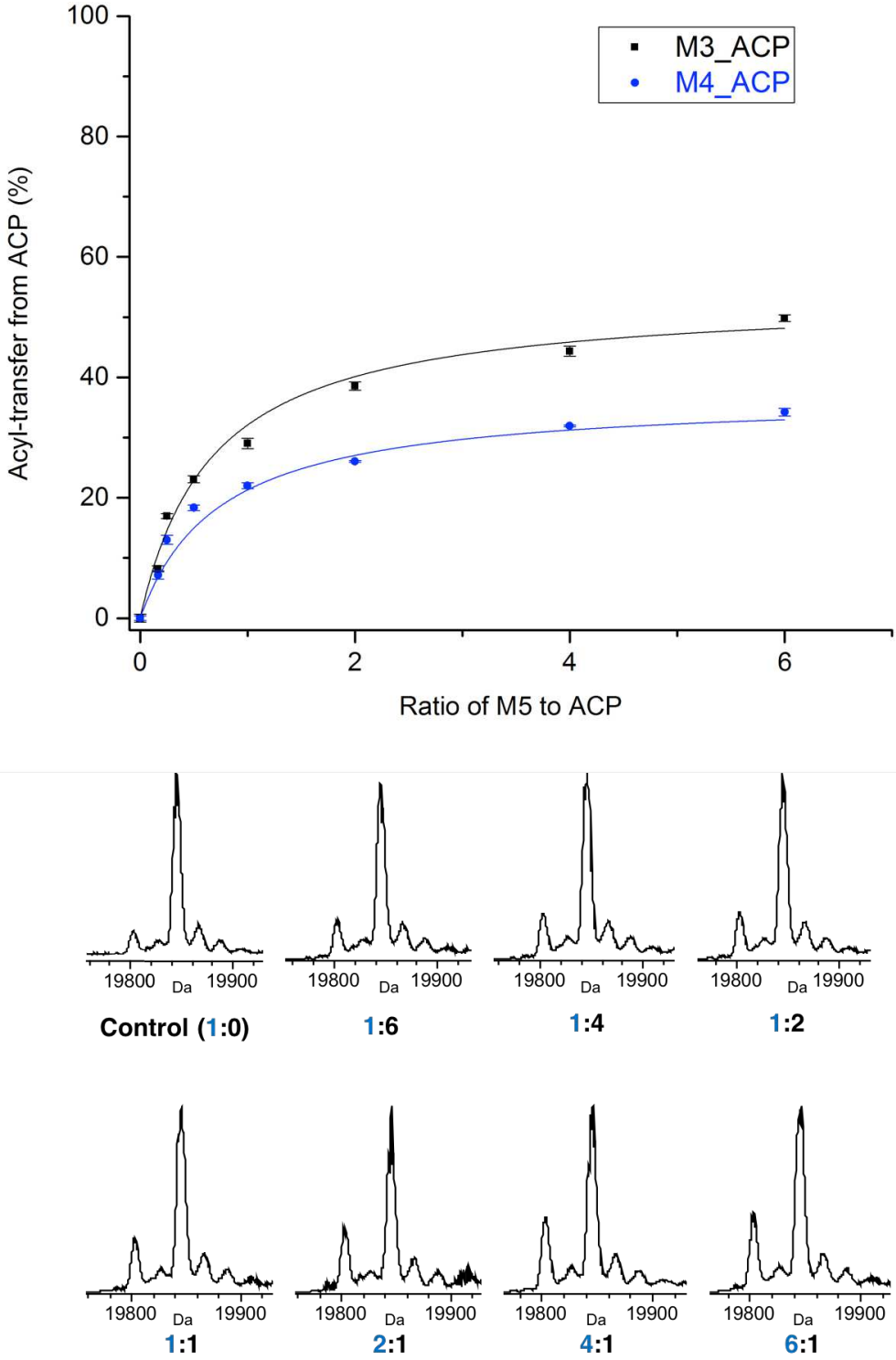


Figure 5.3. *Graphical representation of acyl transfer to M5 for varying equimolar ratios (top) and corresponding deconvoluted spectra for M4_ACP (bottom). Experiments were run in triplicate and errors correspond to one standard deviation.*

5.2 Transfer of 8- and 10-carbon chains from upstream ACPs to module 5

Although the availability of acetyl-CoA makes it convenient to use acetyl groups to monitor transfer, the substrates accepted by M5_KS in the biosynthesis of QMN are 8- and 10-carbon long chains for M3_ACP and M4_ACP, respectively. Transfer of octanoyl and decanoyl substrates from the upstream ACPs to M5 was therefore investigated to confirm the proteins are able to undergo their proposed functions and to determine whether M5 preferentially accepts a specific chain length. An additional advantage is that the increase in mass corresponding to the transfer of larger substrates onto M5 should be detectable by LC/MS, confirming the transfer of acyl groups.

M3_ACP and M4_ACP were loaded with octanoyl- or decanoyl-CoA (mimics of their natural substrates), before incubation with M5 for 12 hours. Controls were run for octanoyl and decanoyl-*holo*-ACPs with M5-C207A, an M5 variant containing a mutation to the KS active-site. No significant levels of conversion to *holo*-ACP were seen in each case, indicating that any substrate transfer seen in the presence of M5 must be to M5_KS's active-site.

Complete conversion to *holo*-ACP can be observed for both ACPs with either substrate in the presence of M5 (Figure 5.4). Furthermore, a corresponding increase in mass was detected for M5 in each case, confirming the transfer of both 8- and 10-carbon chains occurred. The present results do not indicate a clear substrate tolerance, as both reactions proceed to completion. Further kinetic analysis would be required to investigate this further.

Having established that the KS domain of M5 can accept the proposed 8- and 10-carbon chains from M3_ACP and M4_ACP, we sought to more accurately replicate the natural system by allowing module 5 access to both ACPs concurrently. Initially, both ACPs were loaded with octanoyl- or decanoyl-CoA and incubated together in equimolar quantities with twice the amount of M5. LC/MS analysis of 1:1:2 ratio samples (M3_ACP: M4_ACP: M5) confirmed almost complete transfer of both substrates from each ACP and onto M5 (Figure 5.5 (left)). Two M5 variants were observed for each assay, corresponding to both

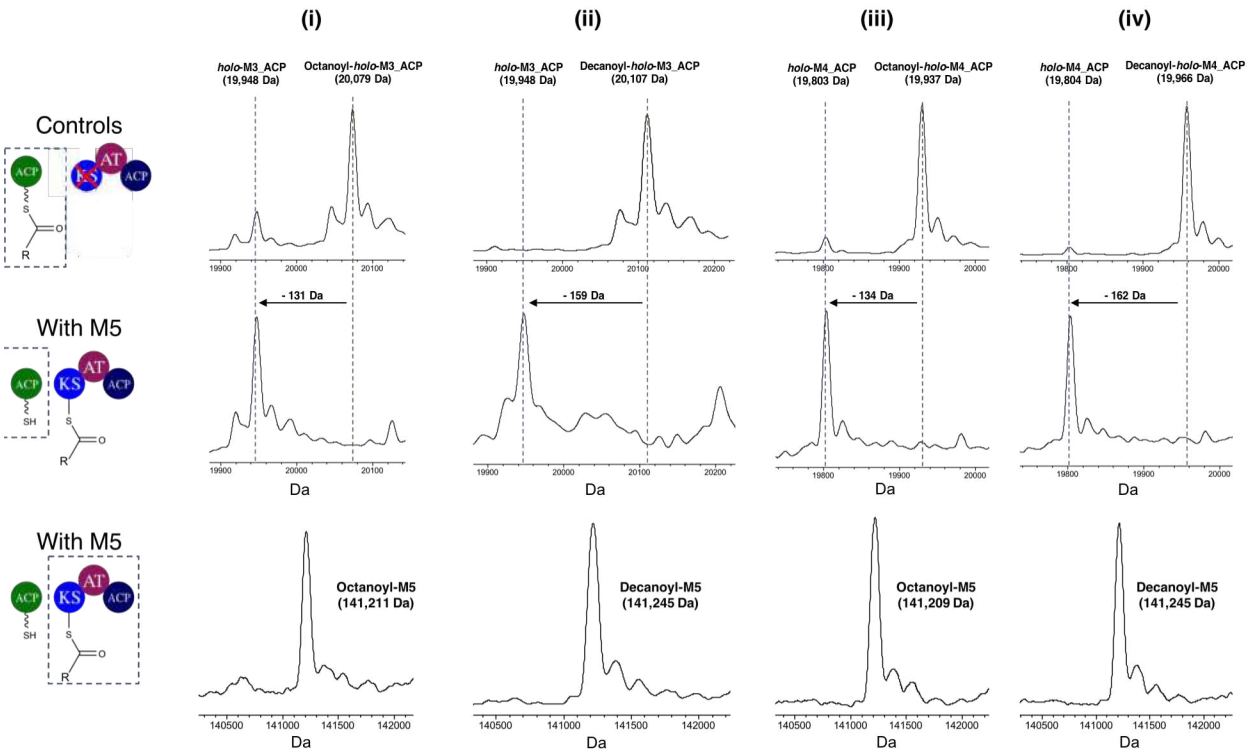


Figure 5.4. *Deconvoluted LC/MS chromatograms showing the transfer of octanoyl and decanoyl substrates from M3_ACP and M4_ACP to M5. The box shows the protein being monitored. ACPs used were **(i)** octanoyl-holo-M3_ACP, **(ii)** decanoyl-holo-M3_ACP, **(iii)** octanoyl-holo-M4_ACP and **(iv)** decanoyl-holo-M4_ACP. Controls were run with M5-C207A.*

octanoyl- and decanoyl-M5. However, as the peaks were not resolved, it was not possible to accurately determine the proportion of each component. The ratio of transferred product was not affected by which ACP was loaded with octanoyl-CoA and which was loaded with decanoyl-CoA.

The experiment was repeated with a 1:1:1 ratio (M3_ACP: M4_ACP: M5) to ensure both reactions could not be completed. Although LC/MS analysis showed a large proportion of M3_ACP was deacylated at the end of the reaction, as would be expected given previous studies on the rate of transfer from M3_ACP and M4_ACP, the difference with the amount of deacylated M4_ACP was relatively small and is not reflected in the ratio of octanoyl-M5 to decanoyl-M5 produced (Figure 5.5 (right)). This is perhaps due to the inaccuracy of ACP concentration.

Although these results correspond with what would be expected given the occurrence of no significant levels of inhibition, time course or kinetic assays are required to confirm whether any inhibition is occurring.

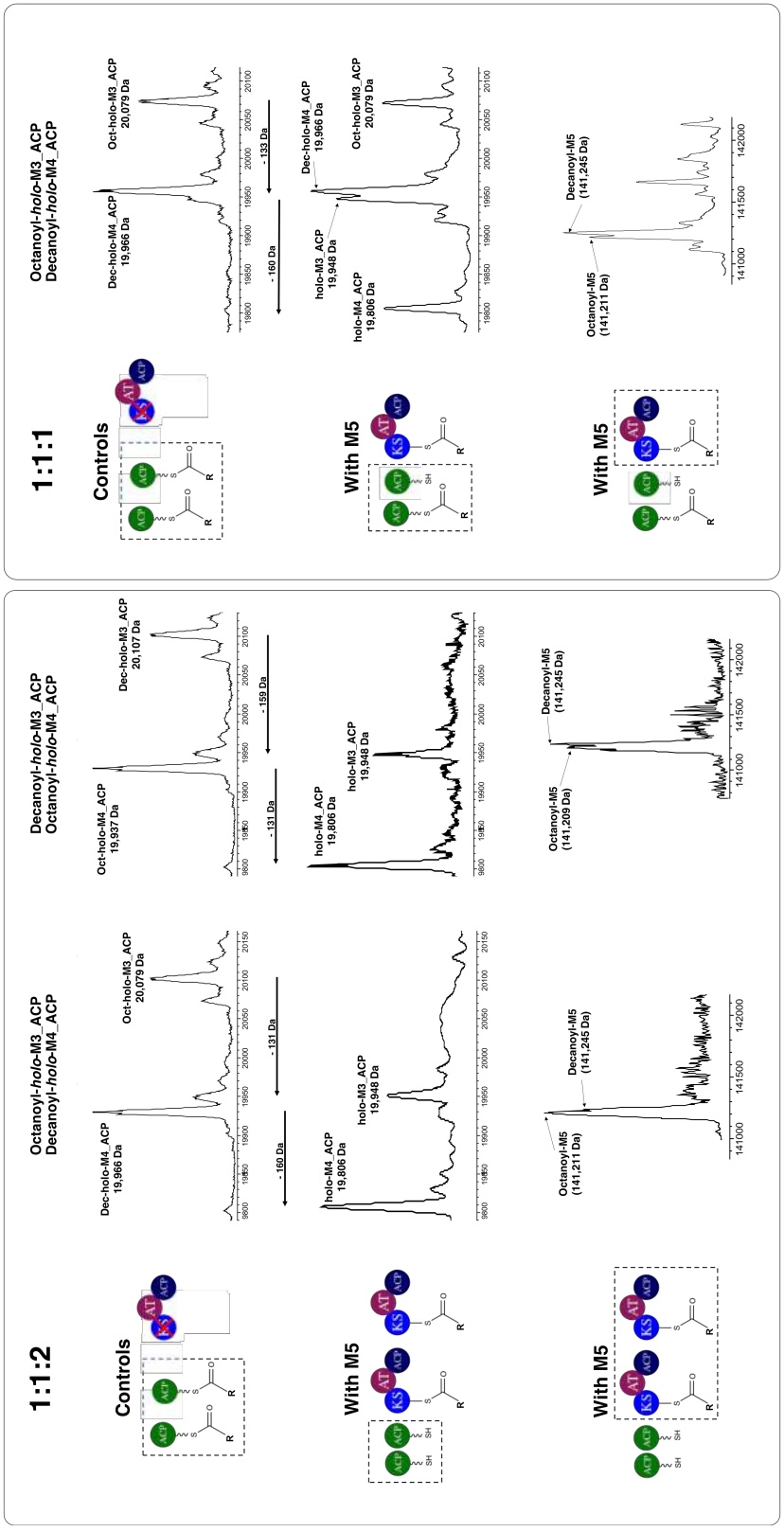


Figure 5.5. *Deconvoluted LC/MS chromatograms of the upstream ACPs and M5 incubated together in a 1:1:1 (left) and a 1:1:2 ratio (right). The box shows the protein being monitored.*

5.3 Intermodular cross-linking of module 5

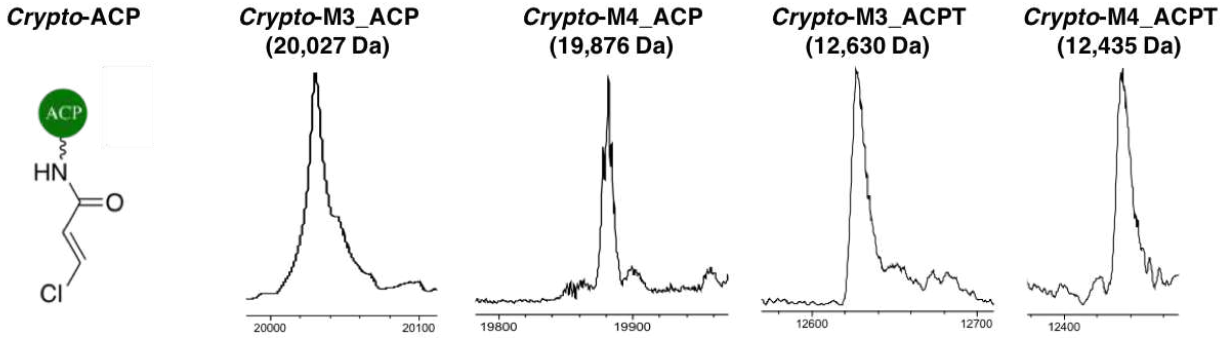


Figure 5.6. *A chemical representation of a crypto-ACP (left) and deconvoluted LC/MS spectra (right) of crypto-M3_ACP (expected mass, 20,019 Da), crypto-M4_ACP (19,879 Da), crypto-M3_ACPT (12,621 Da) and crypto-M4_ACPT (12,429 Da).*

Having established acyl transferase activity of the KS domain, the protein-protein interactions between the upstream ACPs and the KS of M5 were investigated using the previously described chemoenzymatic cross-linking approach (section 3.2). M3_ACP, M3_ACPT, M4_ACP and M4_ACPT were first loaded with the pantetheine chloroacrylamide cross-linker (Figure 5.6) and incubated with M5. SDS-PAGE analysis (Figure 5.7) showed that in both cases a new band appeared, relative to the controls lacking any ACPs. These bands correspond to mass increases of approximately 20 kDa, 15 kDa, 20 kDa and 15 kDa, respectively, indicating successful cross-linking. Incubation with M5-C207A resulted in no observable cross-linking, indicating that the site of cross-linking is specific to the active-site of the domain. Interestingly, incubation with M5-C207A/S656C did not result in cross-linking, suggesting that the upstream ACPs cannot interact with the AT domain in the same way as M5_ACP. The occurrence of cross-linking in the truncated ACPs again suggests the docking domains are not essential in facilitating protein-protein interactions.

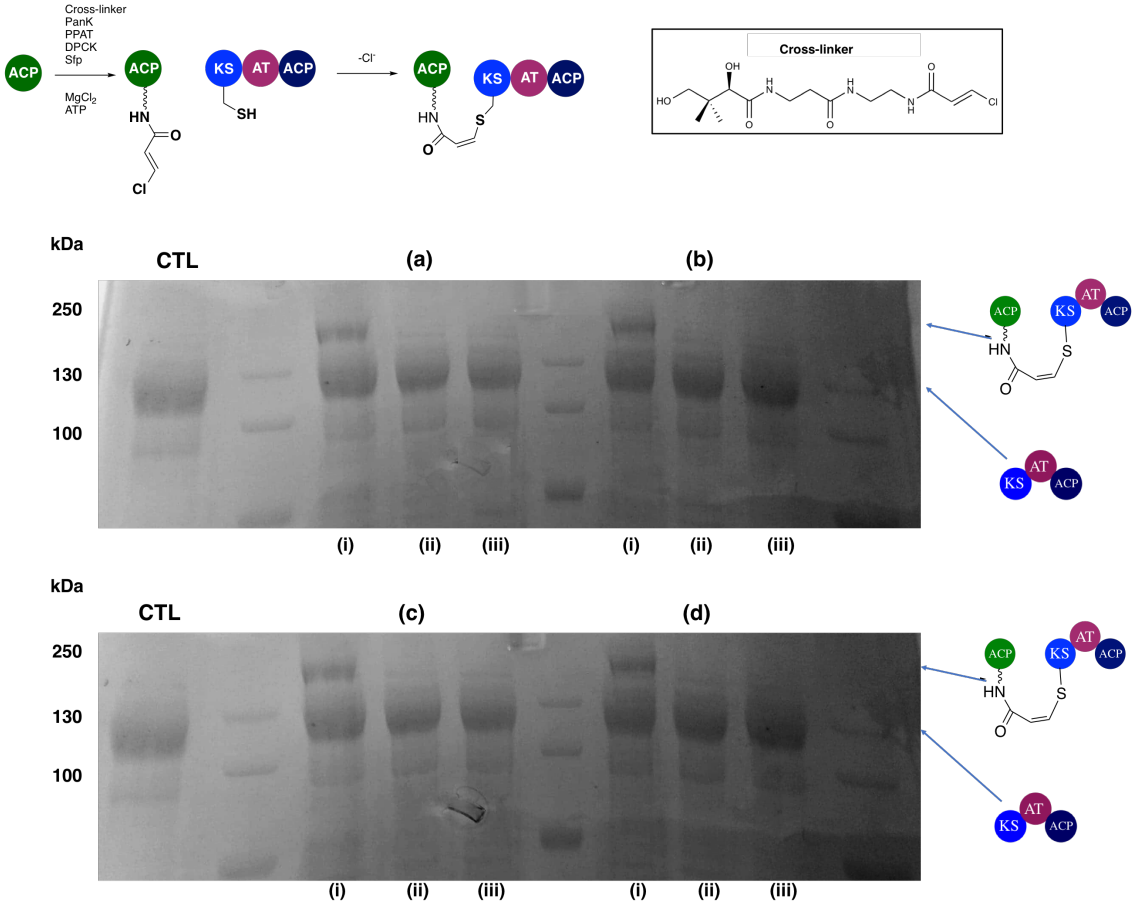


Figure 5.7. *Schematic representation of cross-linking between M5 and an ACP (top) and an SDS-PAGE analysis of cross-linking between M5* ***(i)****, M5-C207A* ***(ii)*** *and M5-C207A/S656C* ***(iii)*** *to M3_ACP* ***(a)****, M3_ACPT* ***(b)****, M4_ACP* ***(c)*** *and M4_ACPT (****d****, bottom). Controls* ***(CTL)*** *correspond to M5 without an ACP.*

5.3.1 Exploring the effects of each region of module 5 on facilitating intermodular interactions

To determine whether all of M5 was essential in facilitating protein-protein interactions with the upstream ACPs, the cross-linking assays were repeated with M5 constructs omitting various domains and structural motifs. SDS-PAGE analysis showed successful cross-linking between the upstream ACPs and M5_KSAT, but no cross-linking to M5_KS or M5_KSATS (Figure 5.8), suggesting the AT-ACP linker region is vital for facilitating upstream interactions with M3_ACP and M4_ACP. As previous work has shown that the constructs lacking a linker region between the AT and ACP domain are unable to function intramodularly and exist as monomers (chapter 4), this highlights the importance of the AT-ACP linker region in both the intra- and intermodular functionality of M5.

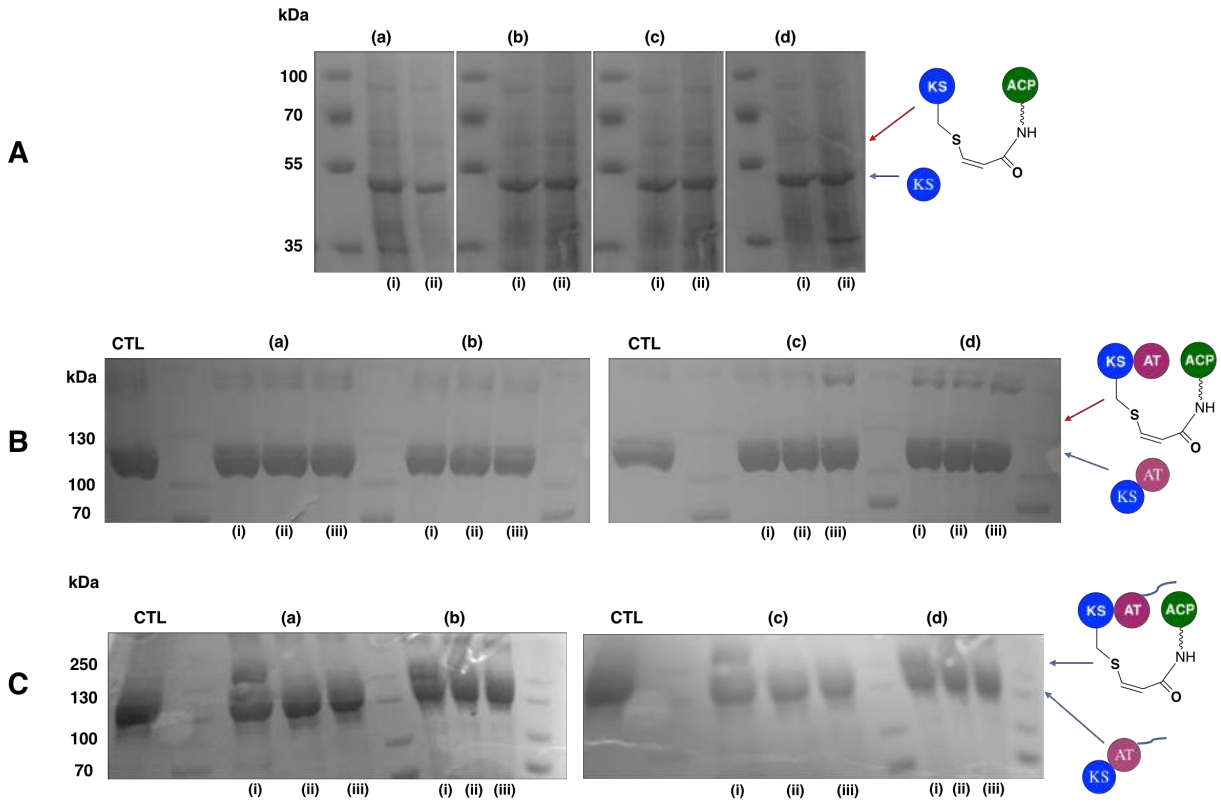


Figure 5.8. *SDS-PAGE analysis of cross-linking of M3_ACP **(a)**, M3_ACPT **(b)**, M4_ACP **(c)** and M4_ACPT **(d)** to the WT **(i)**, C207A **(ii)** and C207A/S656C **(iii)** variants of M5_KS **(A)**, M5_KSATS **(B)** and M5_KSAT **(C)** Controls **(CTL)** correspond to the truncated M5 constructs.*

5.4 Calculating the strength of intersubunit interactions

The cross-linking experiments (section 5.3) give an indication of where protein-protein interactions are occurring, but not about the strength of the interactions. One possible reason for the higher levels of acyl transfer from M3_ACP to M5 may be due to a higher binding affinity between M5 and M3_ACP than is the case for M4_ACP. The stronger binding affinity would increase the likelihood of the upstream ACPs binding to the module and therefore being in close enough proximity to transfer their polyketide substrate.

Attempts to understand the rate of protein-protein interactions using cross-linking proved unsuccessful (Figure 5.9). Surface plasmon resonance (SPR) was therefore used to study the binding and dissociation of M5 with both full-length and truncated versions of the upstream ACPs.

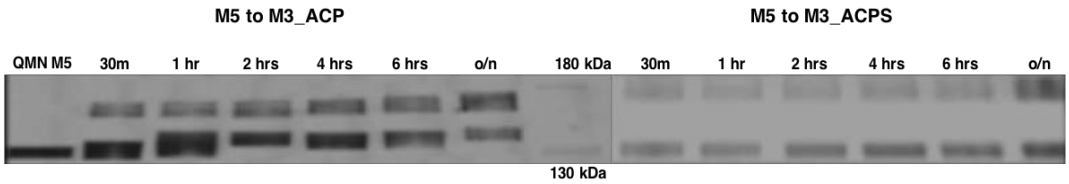


Figure 5.9. *SDS-PAGE analysis of cross-linking at varying time intervals between M5 and M3_ACPT (left) and M3_ACP (right).*

5.4.1 The principles of surface plasmon resonance

When light hits an electrically conducting layer on the interface between materials with a high and a low refractive index (RI) at the resonance angle, a drop in intensity of the reflected light can be detected as a result of an evanescent wave. In SPR, the low RI material is a liquid channel, the high RI material is a glass prism and the gold layer is inside a flow channel of a sensor chip, to which a protein (ligand) may be attached. For this to occur, the surface of the chip must first be activated using 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) and *N*-Hydroxysuccinimide (NHS) before a ligand is immobilised onto the chip through amine coupling and the remaining sites deactivated using ethanolamine-HCl (Figure 5.10). Changes in the aqueous layer close to the chip can be monitored by measuring the RI, allowing confirmation of the ligand binding. The concentration of protein bound to the surface can be varied to adjust the strength of the binding signal, independent of the ligand-analyte affinity.

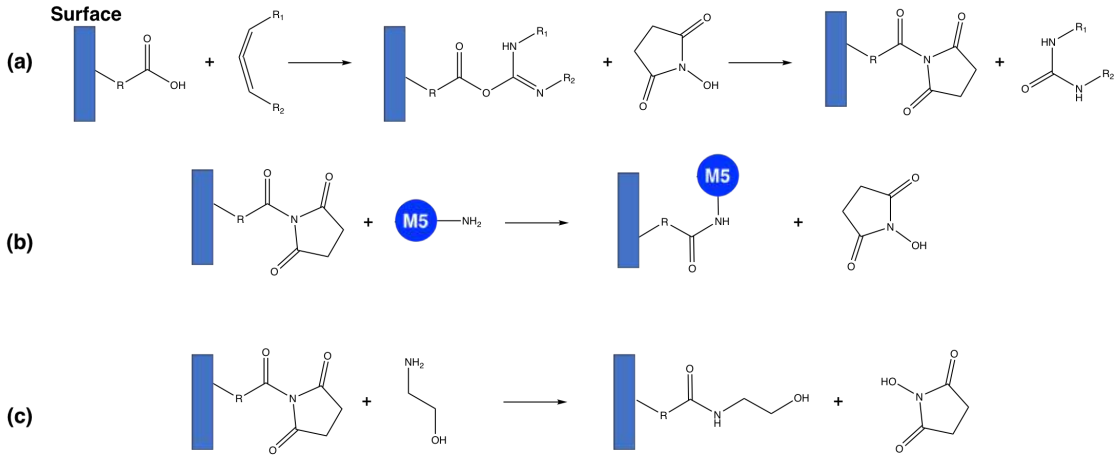


Figure 5.10. *The chemical process behind SPR ligand immobilisation. NHS and EDC activate the surface by the modification of a carboxymethyl group to a N-hydroxysuccinimide ester* ***(a)****. The primary amine of a lysine residue is covalently attached to the surface* ***(b)*** *before the remaining EDC-NHS complexes are deactivated by ethanolamine* ***(c)****.*

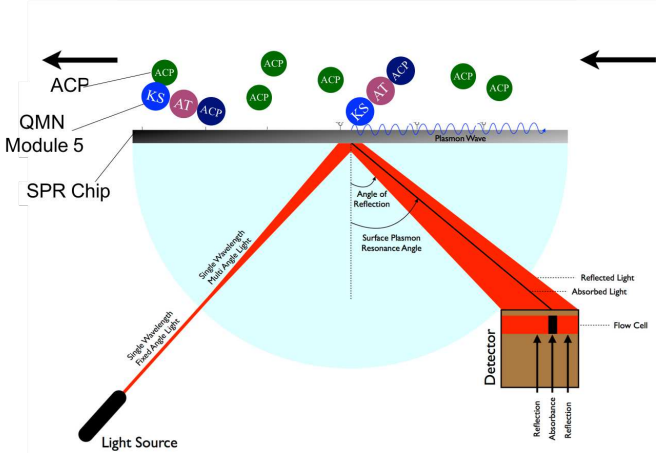


Figure 5.11. *Schematic dipicting the immobilisation of M5 to an SPR chip and interactions with an ACP analyte*[189]

An analyte may then be passed through the channel containing the surface-bound ligand and the response of the surface may be measured (Figure 5.11). The level of response, comparative to an immobilised blank channel, indicates the amount of ligand-analyte binding. A buffer is then run through the channel and dissociation of the analyte from the ligand is measured. Association and dissociation constants may then be calculated by repeating the experiments for a range of concentrations. Kinetic equilibrium association (K_A) and dissociation (K_D) constants are defined as follows:

$$K_A = \frac{k_a}{k_d} = \frac{[AB]}{[A].[B]} \qquad K_D = \frac{k_d}{k_a} = \frac{[A].[B]}{[AB]},$$

where A and B are the ligand and analyte concentrations, respectively, and k_a/k_d are the on- and off-rates. The K_D is the analyte concentration at which half of the binding sites of the surface-bound ligand are occupied. Therefore, a lower K_D value signifies the occurrence of stronger protein-protein interactions.

Immobilisation conditions were tested for a range of sodium acetate buffers (pH 3.5 - 6.3), with pH 6.3 exhibiting the strongest interactions between M5 and the surface (Figure 5.12). pH 6.3 also had the additional advantage of providing the conditions most suitable for amine coupling and was therefore chosen as the immobilisation buffer.

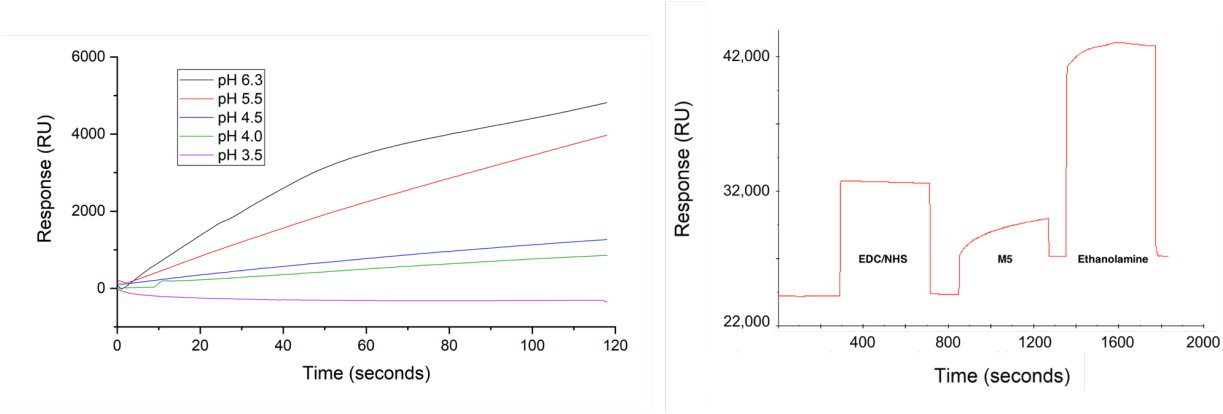


Figure 5.12. *SPR chromatograms of interactions between M5 and the surface in various buffers (left) and successful immobilisation of M5 (right, 3,000 RU).*

5.4.2 Confirming the specificity of the observed interactions

M5 was successfully immobilised (20,000 RU, sodium acetate, pH 6.3) and responses to 20 μM samples of each ACP in its *apo* form were measured (Figure 5.13). Successful regeneration of M5 to its unbound state was observed in each case with a regeneration solution of 10 mM glycine pH 9.0. Maximum RUs were similar for each upstream ACP, with M5_ACP having a slightly lower level of interaction. The responses from a number of *apo*-carrier proteins (CPs) from other systems were also measured, in each case resulting in significantly weaker responses than those observed with the compatible upstream ACPs. AerA is a peptidyl carrier protein (PCP) involved in aeruginosin biosynthesis and contains a short linear motif docking domain. GbnD6 (I) and GbnD6 (II) are both ACPs from the final module of gladiolin PKS. Neither construct contains a docking domain. Gbn constructs were chosen as GbnD6 (I) is proposed to require interactions with a KS, whereas GbnD6 (II) is not.

5.4.3 Elucidating the binding and dissociation constants of module 5 with the upstream ACPs

Binding and dissociation may be more accurately elucidated from curves exhibiting smaller, but still measurable responses ($\sim$ 200 RU). A smaller quantity of M5 was successfully immobilised (3,000 RU, sodium acetate, pH 6.3) and measurable protein-protein interactions were observed with each ACP. Successful regeneration of M5 to its unbound state was observed with a regeneration solution of 10 mM glycine pH 9.0.

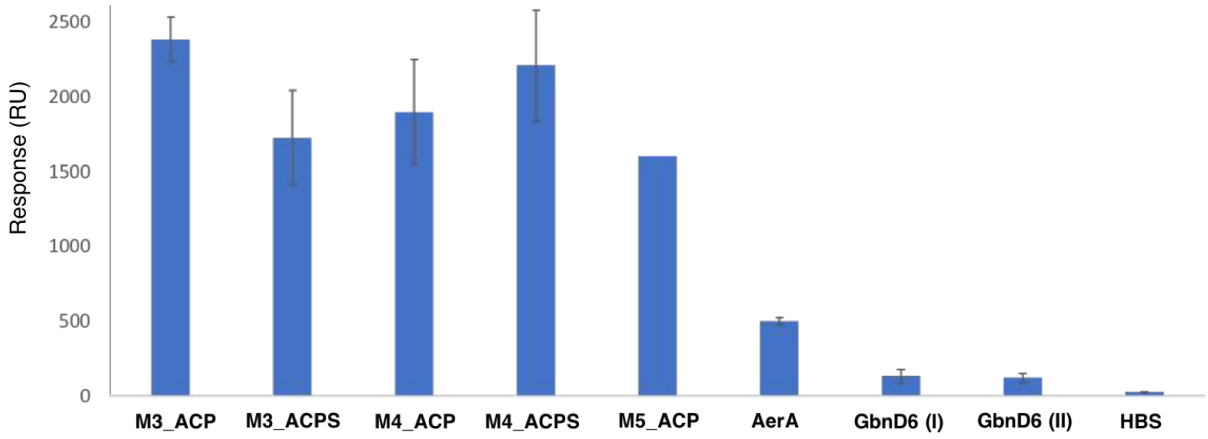


Figure 5.13. *Maximum RU (blank subtracted) for each apo-carrier protein (20 μM) with surface-immobilised M5. Experiments were repeated in triplicate with the displayed errors corresponding to one standard deviation. HBS (HEPES buffered saline) was used as a control.*

Response curves for varying concentrations of each ACP were then acquired (Figure 5.14). Fitting of the data to a number of models resulted in 1:1 binding resulting in the closest fit. This model was used to calculate association and dissociation constants (Table 5.1).

Table 5.1. *Dissociation constants calculated from SPR data of M5 with the upstream ACPs (Figure 5.14).*

Protein	K_D (μM)
M3_ACP	1.3 ± 0.0
M4_ACP	4.0 ± 0.5
M3_ACPT	2.6 ± 0.1
M4_ACPT	4.3 ± 0.7

The data shows that the K_D for M3_ACP is three times lower than for M4_ACP. Furthermore, in the case of M3_ACP, excising the docking domain noticeably weakens the strength of protein-protein interactions ($\sim$ 50 %). This was expected based on previous work suggesting docking domains play a vital role in facilitating intramodular interactions between ACPs and KSs and on our observation that the rate of acyl transfer is faster from M3_ACP than M3_ACPT. Interestingly, in the case of M4_ACP, no statistically significant change in K_D is observed upon removal of the docking domain. This suggests that M4_ACP's docking domain is not effectively facilitating interactions with M5 and explains why M3_ACP was found to have a higher rate of substrate transfer than M4_ACP.

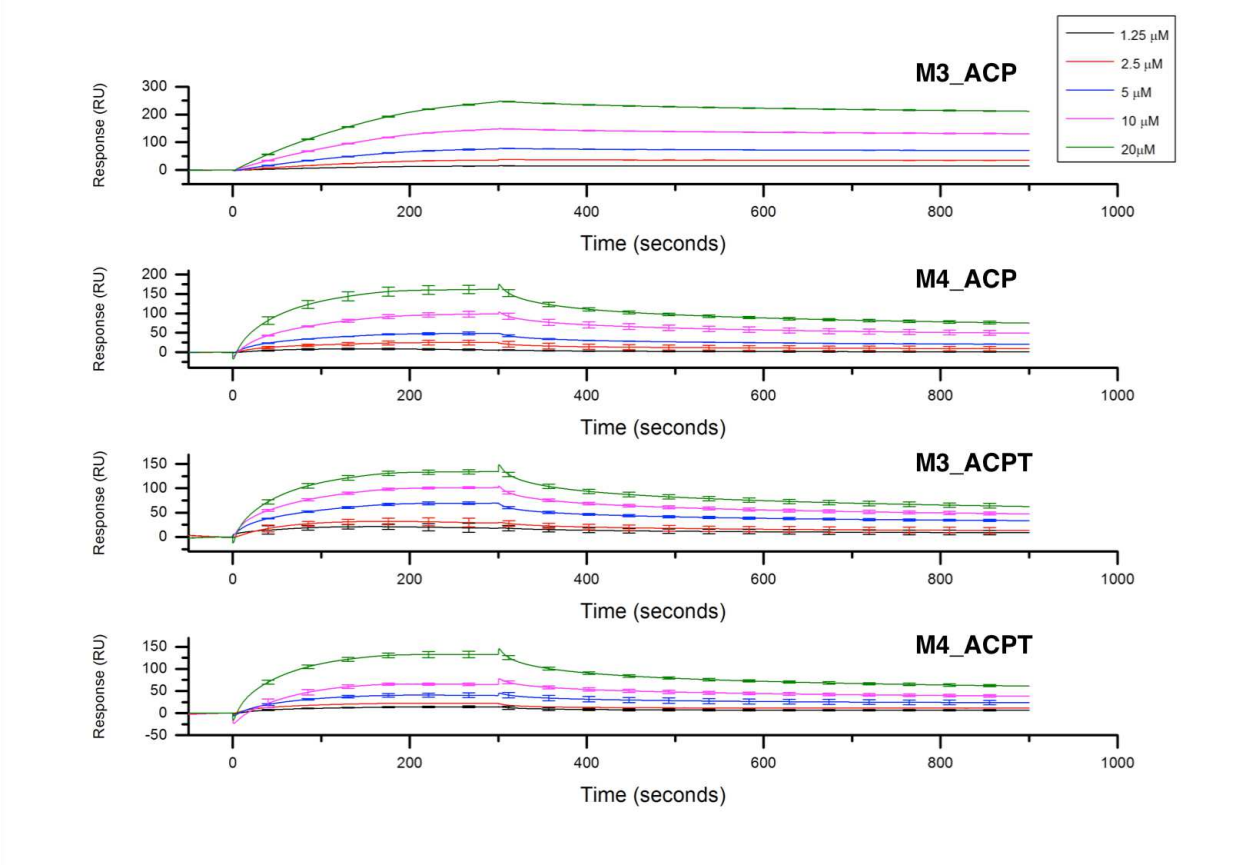


Figure 5.14. *SPR spectra of the upstream ACPs binding to a surface-bound QMN M5. Experiments were run in triplicate with the displayed errors corresponding to one standard deviation.*

5.4.4 Elucidating the binding and dissociation constants of module 5's KS domain with the upstream ACPs

SPR experiments were repeated with M5_KS as the immobilised ligand in order to check whether the other domains in M5 were significantly affecting the intermodular binding and dissociation, and that protein-protein interactions with the upstream ACPs were specific to the KS domain. M5_KS was immobilised (2,000 RU, Figure 5.15) to establish the presence of protein-protein interactions with each ACP. The previously used regeneration conditions were again shown to be suitable.

Kinetic binding experiments were repeated for each ACP (Figure 5.16). A 1:1 binding model was sufficient to calculate kinetic values for M3_ACP and M4_ACP, showing the K_D for M3_ACP to be around four times lower than for M4_ACP (Table 5.2). All math-

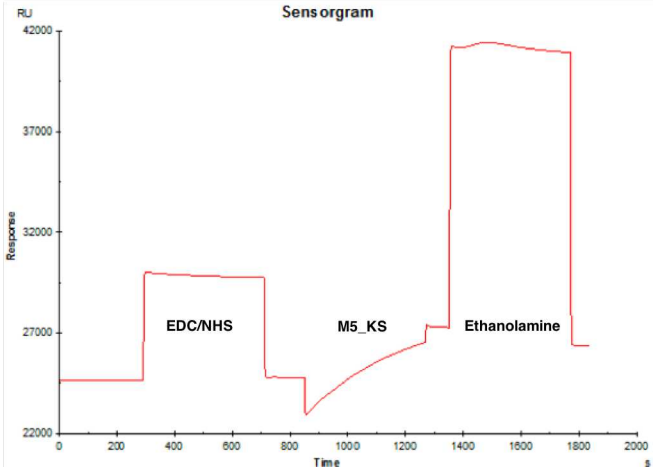


Figure 5.15. *An SPR spectra showing immobilisation of M5_KS (1,800 RU).*

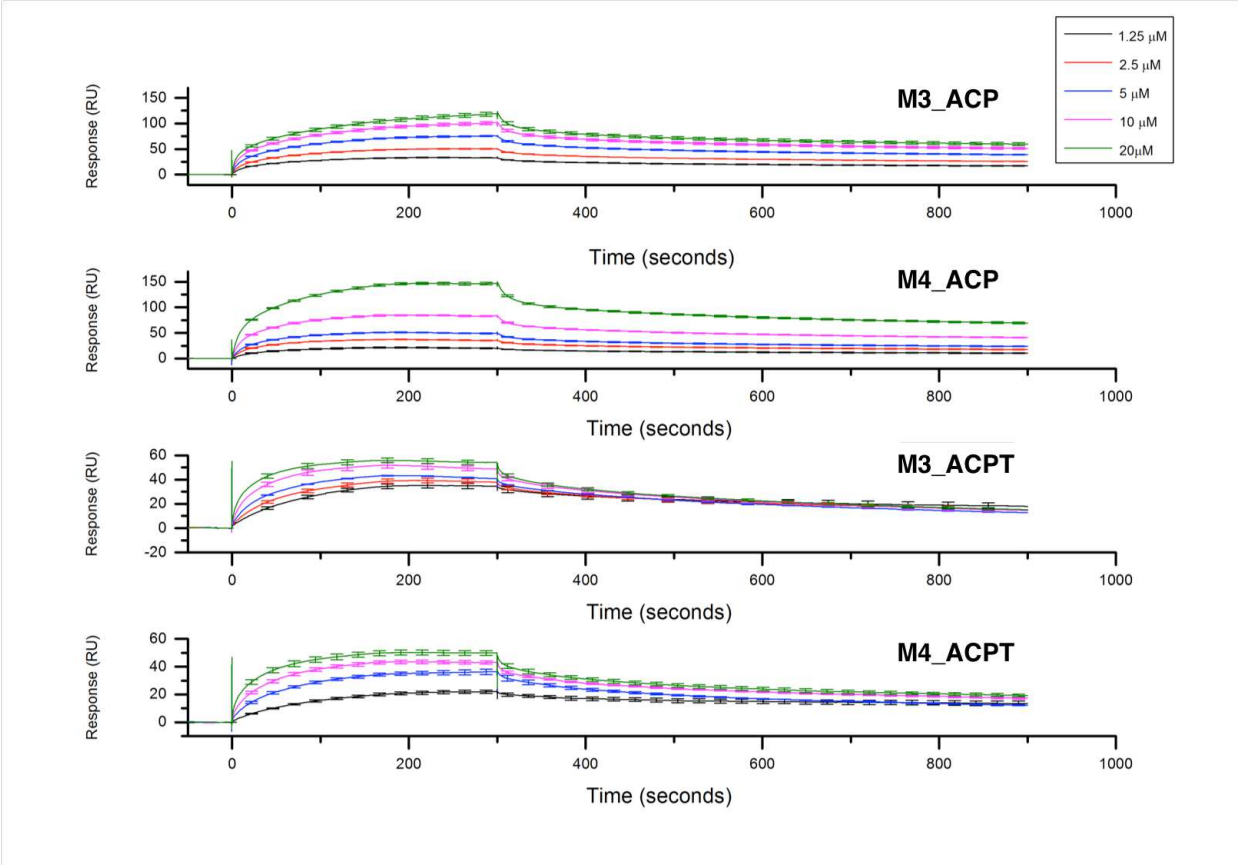


Figure 5.16. *SPR spectra showing interactions between the upstream ACPs and surface-bound M5_KS. Experiments were run in triplicate and the errors correspond to one standard deviation.*

ematical models for the data between M5_KS and the truncated ACPs resulted in large residual errors. The poor fits obtained for the data could perhaps be due to unspecific binding to the KS in areas previously inaccessible due to the presence of an AT domain. Furthermore, the KS construct expressed as a monomer, meaning the residues usually forming an interface between the two KS domains of the homodimeric module are likely

Table 5.2. *Dissociation constants calculated from SPR data of M5_KS with the upstream ACPs (Figure 5.16).*

Protein	K_D (μM)
M3_ACP	0.9 ± 0.0
M4_ACP	3.5 ± 0.1

to be exposed. The increased level of unspecific interactions could interfere with the site-specific interactions seen in section 5.4.3.

No cross-linking had been observed between M5_KS and the upstream ACPs (section 4.4), suggesting any site-specific interactions would be significantly weaker and a high proportion of interactions observed between M5_KS and the upstream ACPs may not be specific to functional interactions. The fact a discernable difference between M3_ACP and M4_ACP was seen could be due to M5_KS's immobilisation to the surface, perhaps through reducing aggregation or presenting the KS domain a more more favourable orientation for interactions.

5.4.5 The effect of malonating module 5 on protein-protein interactions with the upstream ACPs

It had previously been hypothesised that a module conformationally prepares for substrate transfer from an upstream ACP if it is loaded with an extender unit.[190] This would therefore suggest M5 would be more likely to be in a conformation to promote interactions between the KS and the upstream ACPs after the loading of a malonyl-extender unit to M5_ACP. Plans to repeat SPR experiments with immobilised malonyl-*holo*-M5 were made to test this hypothesis. Unfortunately, attempts at malonating M5 (as in section 3.1.2) in the HEPES buffer used for SPR led to instantaneous aggregation upon the addition of $MgCl_2$. M5 was repurified in 50 mM sodium acetate, pH 6.3, and the malonation assay was repeated. Ligand immobilisation was unsuccessful with this sample, with approximately 1 % of the immobilisation levels obtained for similar concentrations of *apo*-M5.

5.5 Structural dynamics of the upstream ACPs

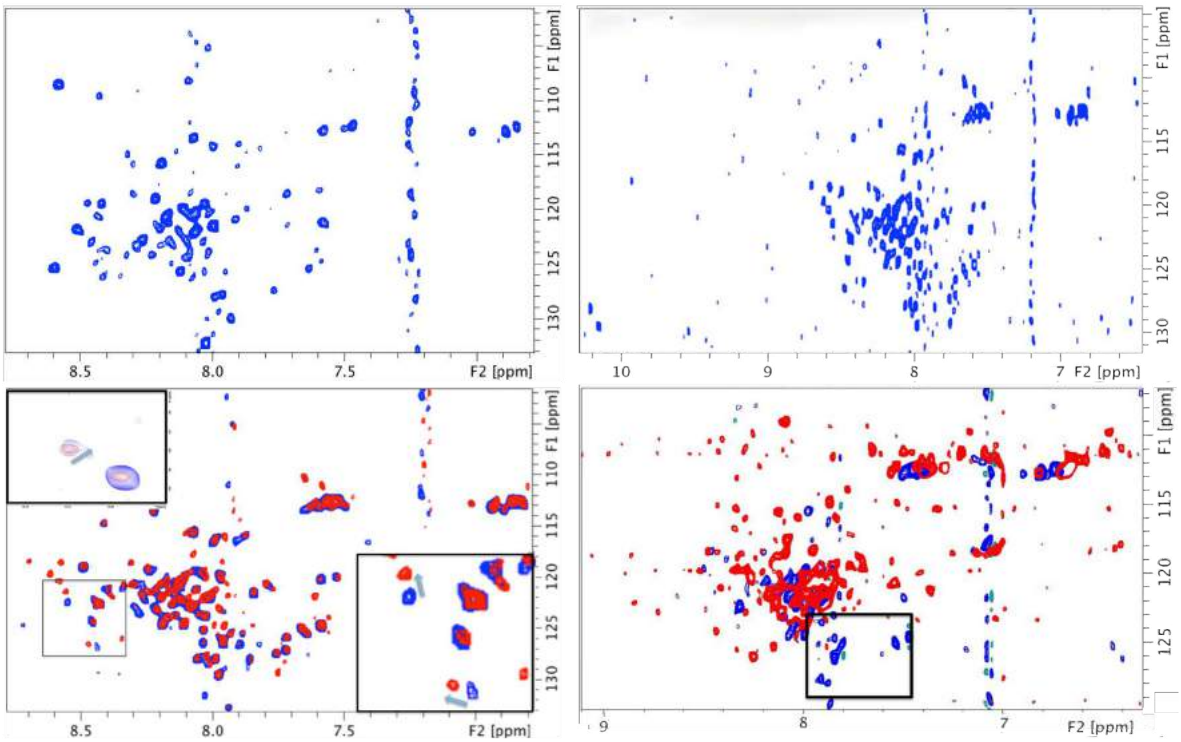


Figure 5.17. *HSQC spectra of [U-^{15}N]-apo-M4_ACP (top left), [U-^{15}N]-apo-M3_ACP (top right), [U-^{15}N]-apo-M3_ACP (bottom left) before and after titration with M5 and [U-^{15}N]-apo-M3_ACP compared with [U-^{15}N]-apo-M3_ACPT (bottom right). The HSQC of [U-^{15}N]-M3_ACP before and after M5 titration displays a magnified insert in the bottom right corner. An insert from a chemical shift of approximately 10 ppm is also shown in the top left. The HSQC comparing M3_ACP and M3_ACPT contains a box that highlights peaks not present in the M3_ACPT spectra.*

Solution NMR was used in an attempt to determine which residues of the upstream ACPs were involved in intermodular interactions with M5. [U-^{15}N]-M3_ACP and [U-^{15}N]-M4_ACP were overproduced and ^{1}H-^{15}N HSQC spectra were acquired for both ACPs (Figure 5.17). Unstructured regions were noted due to the cluster of peaks around 8.1 ppm (F2). However, a reasonable number of peaks were well-defined, suggesting parts of the proteins were well-folded. Titration of M3_ACP with equimolar M5 resulted in a number of observable peak shifts, indicating the occurrence of protein-protein interactions between M3_ACP and M5.

The disordered regions were perhaps due to the constructs containing largely unstructured linker regions between the four conserved helices and the docking domains, giving

the impression that the proteins were partially unfolded. Other reported NMR structures omit the docking domain and linker region, suggesting solution NMR could be a more feasible approach on the ACP constructs containing just the four conserved helices (M3_ACPT and M4_ACPT).

[U-^{15}N]-M3_ACPT was overproduced, purified and a ^{1}H-^{15}N HSQC was acquired in potassium phosphate buffer. A number of peaks disappeared in comparison with full-length M3_ACP and the dispersion of peaks significantly improved, suggesting removal of disordered regions. The area highlighted contains numerous peaks only visible in the spectrum for M3_ACP, likely to correspond to residues of the docking domain. Some of the missing peaks were previously shown to shift upon titration of M5 and some were unaffected. In addition, a number of peaks observed to shift upon titration of M3_ACP with M5 were still present, suggesting residues in the docking domain and the four conserved helices interacted with M5.

5.6 Structural elucidation of the docking domains

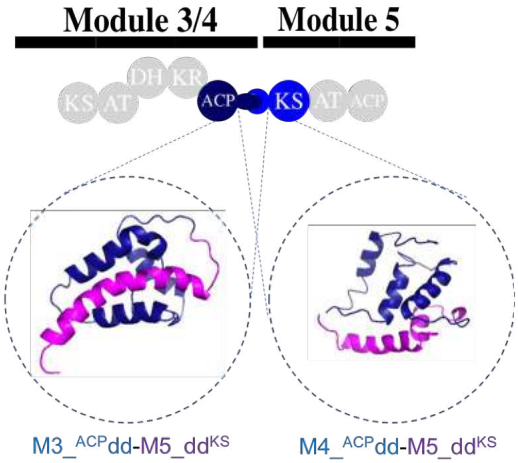


Figure 5.18. *Homology models of the ACP docking domains from M3 and M4 fused to the KS docking domain of M5 (112 and 111 residues, respectively).*

Structural elucidation of the upstream ACP's docking domains in complex with the KS docking domain of M5 (M5_dd^{KS}) could help understand how they interact together (Figure 5.18). NMR has the ability to determine the spatial placement of each residue, thereby

determining how electrostatic interactions form between the helices. Mutation of key residues could also allow the elucidation of which residues are essential for facilitating and maintaining interactions.

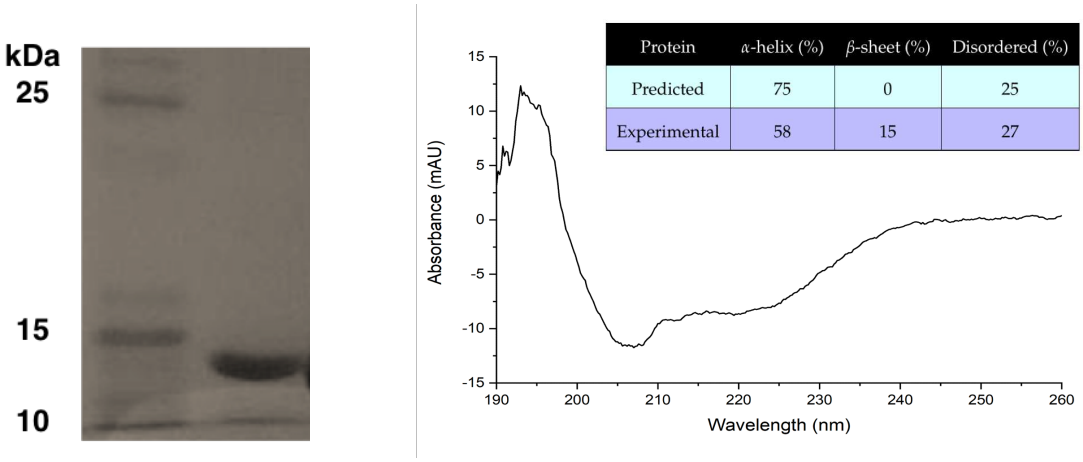


Figure 5.19. *SDS-PAGE analysis of His-tagged M4_^{ACP}dd-M5_dd^{KS} (15,790 Da, left) and a CD spectra (right).*

A fused complex between M4's ACP docking domain (M4_^{ACP}dd) and M5_dd^{KS} was overproduced and purified, before structural integrity was confirmed by CD (Figure 5.19). No linker region was present, with the KSdd starting immediately after the end of the ACPdd. Production of $^{13}C/^{15}N$ labelled protein could allow structural elucidation of the interacting docking domains, helping to understand how the electrostatic interactions play a role in intermodular interactions. A similar approach could be applied to a fused M3/M5 docking domain construct.

5.7 Conclusion

It has been established through a series of transfer assays that both upstream ACPs are able to transfer various substrates and that M3_ACP more efficiently transfers substrates to M5 than M4_ACP. Acyl transfer assays with truncated constructs omitting a docking domain resulted in transfer being seen for both ACPs. M3_ACPT had a significantly slower rate of transfer than M3_ACP. However, no significant difference in the rate of transfer was seen between M4_ACP and M4_ACPT. Interestingly, the rate of transfer for both M4 constructs was comparable to the rate seen for M3_ACPT.

SPR data indicated this was due to M3_ACP's docking domain more efficiently facilitating protein-protein interactions than M4_ACP's docking domain and therefore preferentially transferring substrates to M5. Such a bias in module preference is likely to be due to M3_ACP being able to transfer a substrate to both M4_KS and M5_KS. Efficient QMN biosynthesis requires an equal number of precursors **1** and **2** to be produced. As nature usually produces extremely efficient pathways, it is reasonable to hypothesise that approximately 50% of the substrates bypass M4.

Structural elucidation of M3_ACPT and M4_ACPT would be completed by acquiring HSQC, cbcaconh and cbcanh spectra for both constructs. Acquisition of HSQC spectra for M3_ACP and M4_ACP and both ACPs titrated with M5 would then allow the identification of which residues of the ACPs' four conserved helices interact with M5. Structural elucidation of the fused docking domain constructs would also provide detailed information about molecular interactions between the docking domains.

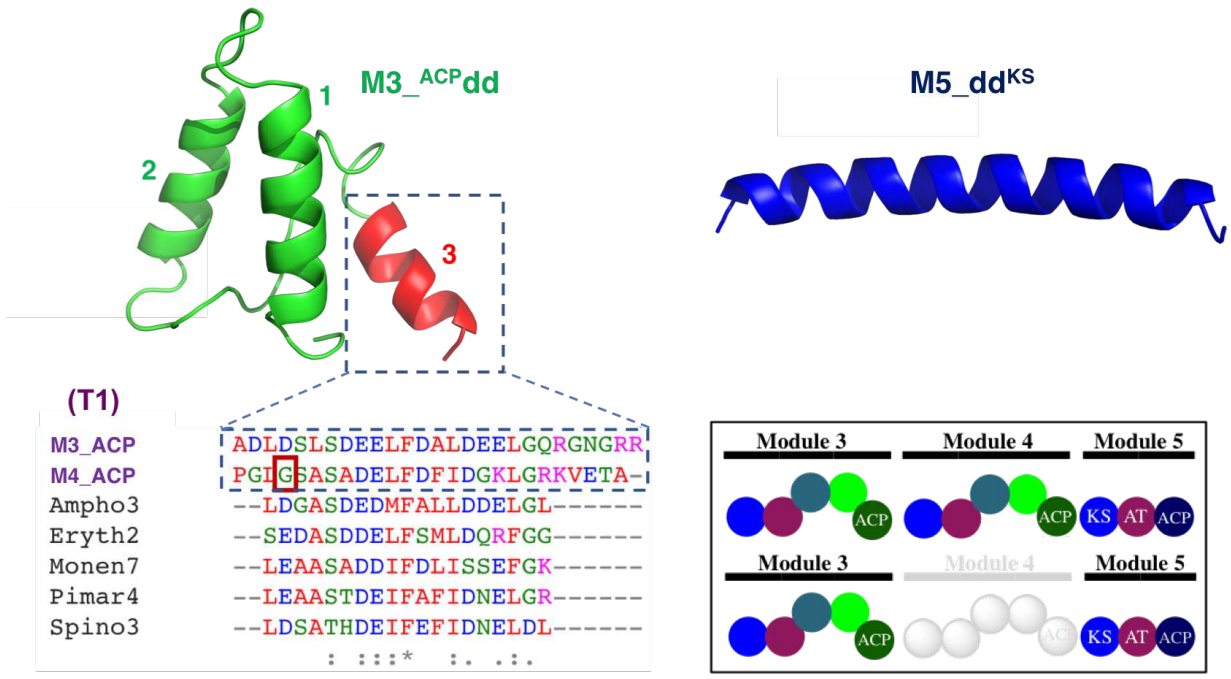


Figure 5.20. *A homology model of M3_ACP's docking domain (top left), a homology model of the M5_KS docking domain (top right), sequence comparison of the final helix of the ACPs with other Class 1 Type 1 docking domains (bottom left) and a schematic representation of the domains involved in QMN module skipping (bottom right).*

Sequence alignment of M3_ACP and M4_ACP with other known T1 docking domains (Figure 5.20) showed the widely conserved negative charged aspartic or glutamic acid residue at the start of the third helix was replaced by a glycine in M4_ACP. The electrostatic pattern of the third helix is widely reported to be essential in facilitating

docking interactions with KS docking domains.[131] Therefore, the absence of a negatively charged residue in M4_ACP could interfere with the electrostatic pattern and disrupt the occurrence of protein-protein interactions between the two docking domains. Mutation of M4_ACPdd's glycine residue (G126) to either an aspartic or glutamic acid could potentially create a mutant with greater affinity for protein-protein interactions with M5's KS domain, hence providing evidence to support this hypothesis. In addition, mutation of the aspartate in M3_ACP would allow further studies to determine if its ability to facilitate interactions with M5 had decreased, demonstrating the importance of this negative residue in intramodular interactions.

Finally, to gain a complete picture of how the QMN module skipping mechanism works, interactions between M3_ACP and M4 must also be studied. Many of the experiments described in this chapter could be repeated to probe the acyl transfer and protein-protein interactions between module 3 and 4 and how this compares to their interactions with M5.

When combined together, these approaches would provide the site-specific interactions of all the domains involved in the two different pathways and create a better understanding of the module skipping mechanism present in QMN PKS. This information will prove invaluable in the possible creation of synthetic module skipping PKSs with the ability to produce multiple natural products.

Chapter 6

Conclusions and future work

6.1 Activity and substrate tolerance of quartromicin module 5

A series of assays, monitored by intact ESI mass spectrometry, showed QMN module 5 is capable of accepting a substrate from the ACP domains of either module 3 or module 4 and of catalysing chain extension of the substrates with a malonyl-CoA extender unit, thus confirming the predicted role of module 5 in the biosynthesis of quartromicin. Successful transfer of 2, 8 and 10-carbon chain substrate mimics from both M3_ACP and M4_ACP to the KS domain, as well as successful chain extension of a wide range of substrate analogues (including those with shorter, longer, saturated and branched carbon chains) with a malonyl-CoA extender unit indicated that the substrate tolerance of module 5 is not a key factor in governing pathway fidelity. M5 was, however, unable to catalyse chain extension using methylmalonyl-CoA. Direct loading of a ppant-methylmalonyl group to the ACP and incubation with an octanoyl substrate led to the detection of extended product, showing that the extender unit specificity of M5 is ensured by the AT domain. These results indicated that the pathway's fidelity is more likely ensured by the specificity of protein-protein interactions between module 5 and the ACP domains of modules 3 and 4.

Although attempts were made to investigate the structure of module 5 by NMR, X-ray crystallography, carbene footprinting and cryo-EM, none were particularly successful due to the propensity of module 5 to aggregate. Future work could involve repetition of these studies with homologous modules from the potential quartromicin producing *Amycolatopsis albispora* to help understand the architecture and mechanistic rearrangements of each module throughout its catalytic cycle.

6.2 Importance of docking domains in facilitating protein-protein interactions between module 5 and the ACP domains of modules 3 and 4

In addition to confirming the ability of M5 to catalyse acyl-transfer from M3_ACP and M4_ACP, the presence of corresponding protein-protein interactions was confirmed by intermodular cross-linking. These experiments indicated that whilst the upstream ACPs can interact with the KS of M5, they can not interact with the AT domain. Analysis of acyl-transfer rates showed that M3_ACP more efficiently transfers a substrate to M5 than M4_ACP. Removal of the docking domains had no significant effect on the rate of transfer from M4_ACP, whereas a significant reduction was noted in the case of M3_ACP.

SPR analysis with chip-immobilised M5 confirmed the occurrence of protein-protein interactions between M5 and both full length and truncated M3_ACP and M4_ACP. Kinetic studies suggested that M3_ACP had a lower K_D than M4_ACP (1.3 v 2.6 μM). In comparison, truncated versions had K_D values of 4.0 and 4.3 μM for M3_ACPT and M4_ACPT, respectively, suggesting that M3's ACP docking domain more efficiently facilitates protein-protein interactions than M4's. This is perhaps because a substrate attached to M4 may only be transferred to M5, whereas a substrate attached to M3 has the potential to be transferred to M4 or M5.

Whilst both ACP docking domains are structurally similar, the reduced efficiency of M4's could be due to the presence of a glycine residue, in place of a widely conserved negatively charged residue, interfering with the electrostatic pattern required for facilitating interactions. Mutation of M3_ACP's conserved negative residue to a glycine and M4_ACP's glycine to a negative residue, followed by the repetition of the *in vitro* assays and SPR experiments would test the aforementioned hypothesis. However, it is important to note that the creation of mutated constructs has previously had undesirable effects in other fused docking domain systems.[124]

Initial solution state NMR experiments indicated that residues within both the main 4-helix bundle and the docking domain of M3_ACP were involved in protein-protein interactions with M5. This work could be completed by carrying out more extensive titration experiments following HSQC assignment of both M3_ACP and M4_ACP. Furthermore, structural elucidation of the fused docking domain complexes (section 5.6) would show how the docking domains interact on an atomic-level.

A canonical understanding of docking domain interactions could lead to the production of synthetic docking domains to facilitate specific intersubunit interactions. Understanding how numerous precursors can be created by a pathway and how quantities are controlled is vital in creating novel PKS pathways that produce numerous precursors.

6.3 Identification of a dimerisation motif essential for modular dimerisation and catalytic functionality

A loss of dimerisation was noted when the AT-ACP linker region was omitted from M5 constructs, leading to the hypothesis that this region contains a dimerisation motif. Homology models predicted the region to be mostly structured and to contain an extended β-sheet and a partial Rossmann fold. Comparison with other final modules of spirotetronate PKSs was carried out as they are all proposed to end with a stand-alone non-reductive module. The majority of modules contained a significantly shorter unstructured AT-ACP linker region. However, ChlA6 and KijS5 were identified as containing long, highly structured linker regions with similarities to M5. Sequence analysis indicated they are likely to contain the remnants of a KR domain. In addition to the expected large interaction interface between the KS domains, 2-D cryo-EM reconstructions of M5 also depicted what seemed to be a contact region at the other end of the complex. This is likely to be formed by the dimerisation motif present in the linker region.

Analytical gel filtration of both C-terminal and N-terminal truncated M5 constructs indicated that a dimerisation motif existed, with a loss of dimerisation noted between residues 1126 - 1160. A chemoenzymatic cross-linking approach demonstrated this dimerisation motif, which we have termed a KR^D domain, was essential for facilitating interactions between both the upstream ACPs and M5's ACP domain, thus ensuring modular functionality.

Successful crystallisation and subsequent structural elucidation of the linker region would help explain the region's role in module dimerisation. Although initial crystal trials proved unsuccessful, various lengths of construct and a wider range of conditions may still allow crystals to be obtained. Acquiring atomic-level resolution information on the dimerisation motif would highlight which residues are essential in the formation of a dimerisation bridge, as well as understanding the structure of the formed bridge. This could potentially allow the identification of similar motifs in other PKS systems. Similar structural studies could be carried out on ChlA6, KijS5, the M5 homologue from *Amycolatopsis albispora* and any other systems found to contain dimerisation motifs, to provide a deeper understanding of how these motifs function.

Chapter 7

Experimental Procedures and Methods

7.1 Materials and Instruments

7.1.1 Chemicals and reagents

All chemicals and reagents used in this thesis were of molecular biology grade and purchased from the stated suppliers. All guidelines were closely followed. dH_2O was prepared by autoclaving H_2O (121°C, 20 minutes) and used to dilute all chemicals to the required working concentration (unless otherwise stated).

7.1.2 Stock solutions

Table 7.1. *The solvent for all stock solutions was dH_2O. All solutions were sterilised by autoclave.*

Stock solution	Concentration	**Manufacturer**
Trisaminomethane-hydrochloride (Tris-HCl)	0.2 M	Thermo Fisher
Sodium Chloride (NaCl)	0.1 M	Thermo Fisher
Imidazole	1.0 M	Sigma-Aldrich
Ethylenediaminetetraacetic acid (EDTA)	0.1 M	Sigma-Aldrich
Isopropyl β-D-1-thiogalactopyranoside (IPTG)	1.0 M	VWR Chemicals
Magnesium chloride ($MgCl_2$)	0.1 M	Thermo Fisher
Calcium chloride ($CaCl_2$)	0.1 M	Thermo Fisher
Potassium phosphate (KH_2PO_4)	0.1 M	Scientific Laboratory Supplies

10 % ammonium persulfate (APS, Thermo Fisher), 10 % sodium dodecyl sulfate (SDS, Sigma-Aldrich), 100 mg/mL ampicillin (Sigma-Aldrich), 50 % glycerol (Sigma-Aldrich) and 70 % ethanol are made up by dilution with dH_2O and filtered using MF-Millipore^{TM} membrane filters (0.2 μm, Merck Millipore). All other buffers were prepared as described in Table 7.1.

7.1.3 Instruments and equipment

PCR reactions were performed using the Eppendorf Mastercycler® nexus GX2 thermocycler. Negative-stain images were acquired on a Jeol 2010F transmission electron microscope. LC-QTOF was performed using the maXis II™ (Bruker) and MALDI using an Autoflex speed MALDI-TOF/TOF spectrometer (Bruker). An AmaZon X (Bruker) was also used for LC/MS. Circular dichroism spectra were acquired using a JASCO 1500 instrument and HSQCs were acquired using a Bruker Avance II 700 MHz spectrometer. Surface plasmon resonance data was acquired on Biacore T2000 and Biacore T200 instruments (GE Healthcare Life Sciences).

Other instruments used: Astell compact top autoclave, New Brunswick Scientific Innova® 44 refrigerated incubator shaker, ÄKTA Pure 25 L (GE Healthcare Life Sciences) fast liquid protein chromatography (FPLC) system, Thermo Electron Corporation Heraeus® Function Line incubator, an Eppendorf® Centrifuge 5804 R, HI-VAC® vacuum valve (Sigma-Aldrich) and a VirTis BenchTop™ K lyopholiser.

7.1.4 *Escherichia coli* strains

Table 7.2. *Strains of Escherichia coli used in this thesis.*

Strain	Manufacturer	Description	Genotype
One Shot® TOP 10™	Invitrogen	Plasmid holding	F-mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 ΔlacX74 nupG recA1 araD139 Δ(ara-leu)7697 galE15 galK16 rpsL(StrR) endA1 λ-
One Shot® BL21 Star™ (DE3)	Invitrogen	Protein overproduction	B F-ompT gal dcm lon hsdSB(rB?mB?) λ(DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]) [malB+]K-12(λS)
NiCo21 (DE3)	New England Biolabs	Protein overproduction	can::CBD fhuA2 [lon] ompT gal (λ DE3) [dcm] arnA::CBD slyD::CBD glmS6Ala ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5
OverExpress C43 (DE3)	Lucigen	Protein overproduction	F-ompT gal dcm hsdSB(rB- mB-)(DE3)
Rosetta™ 2 (DE3) Singles™	Novagen	Protein overproduction	F-ompT gal dcm lon? hsdSB(rB-mB-) λ(DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]) [malB+]K-12(λS) pLysSRARE[T7p20 ileX argU thrU tyrU glyT thrT argW metT leuW proL orip15A](CmR)

7.1.5 Buffers and gels

All buffers used with an FPLC were vacuum filtered using a sterile 500 mL Steritop-GP 45 mm threaded bottle top filter (0.22 μm pore size, polyethersulfone (PES) Express PLUS membrane, Merck Millipore).

Wash buffer

- 20 mM Tris-HCl
- 100 mM NaCl
- 20 mM Imidazole
- Add dH_2O up to 1 L
- pH 8.0

Elution buffer

- 20 mM Tris-HCl
- 100 mM NaCl
- 500 mM Imidazole
- Add dH_2O up to 1 L
- pH 8.0

Benchtop elution buffers

- 20 mM Tris-HCl
- 100 mM NaCl
- 50, 100, 200 and 300 mM Imidazole
- Add dH_2O up to 1 L
- pH 8.0

Storage buffer

- 20 mM Tris-HCl
- 100 mM NaCl
- Add dH_2O up to 1 L

- pH 8.0

Anion exchange (AEX) buffers

- 20 mM Tris-HCl
- 200, 300, 400, 500 mM and 1 M NaCl
- Add dH_2O up to 1 L
- pH 8.0

10x SDS-PAGE (protein gel) running buffer

- 30 g Tris-HCl
- 144 g Glycine
- 10 g Sodium dodecyl sulfate
- Add dH_2O up to 1 L

6x SDS-PAGE loading dye

- 2 mL 1 M Tris-HCl pH 6.8
- 2 mL Glycerol
- 0.4 mL β-mercaptoethanol
- 0.02 g Bromophenol blue
- 0.4 g Sodium dodecyl sulfate
- 2 mL dH_2O

IMAC column stripping buffer

- 10 mM Tris-HCl
- 100 mM NaCl
- 100 mM EDTA
- Add dH_2O up to 1 L

0.1 M Nickel(II) sulfate ($NiSO_4$)

- 1.55 g $NiSO_4$
- 100 mL dH_2O

0.1 M Cobalt(II) sulfate ($CoSO_4$)

- 1.55 g $CoSO_4$
- 100 mL dH_2O

10x Tris-borate-EDTA (TBE) elecrophoresis running buffer

- 108 g Tris-HCl
- 55 g Boric acid
- 20 mL 1 M EDTA

6x gel loading dye

- 2 mL Glycerol
- 0.02 g Xylene cyanol FF
- 0.02 g Bromophenol blue
- 4 mL dH_2O

1 % agarose gel

- 1 g agarose
- 10 mL 10x TBE buffer
- 90 mL dH_2O
- Dissolved in microwave and cooled
- Add 5 μL of GelRed^{TM} Nucleic Acid Stain (Biotium)
- Set in gel plate with gel comb (20 minutes)

Circular dichroism (CD) buffer

- 10 mM potassium phosphate
- 20 mM NaCl

Nuclear magnetic resonance (NMR) buffer

- 2.5 mL of 0.1 M Potassium phosphate
- 5 μL of 0.1 M DSS
- 250 μL D_2O

- Up dH_2O up to 5 mL
- pH 6.0, 6.5 or 7.0

Liquid Chromatorgraphy/ Mass Spectrometry (LC/MS) buffer A

- dH_2O (99.9 %)
- 0.1 % Formic acid

LC/MS buffer B

- Acetonitrile (99.9 %)
- 0.1 % Formic acid

7.1.6 Culture media

Lysogeny Broth (LB)

- 25 g Lysogeny Broth, Miller (Thermo Fisher)
- Add dH_2O up to 1 L and autoclave
- 1 mL Ampicillin (100 mg/mL)

Lysogeny Broth Agar

- 25 g Lysogeny Broth (Thermo Fisher)
- 15 g Bacto agar (Sigma-Aldrich)
- Add dH_2O up to 1 L and autoclave
- 1 mL Ampicillin (100 mg/mL)

5 x M9 salt solution

- 32 g $Na_2HPO_4 \cdot 7\,H_2O$
- 6 g KH_2PO_4
- 2.5 g NaCl
- Add dH_2O up to 500 mL and autoclave

M9 minimal media

- 200 mL 5 x M9 salt solution

- 2 mL 1 M $MgSO_4$
- 100 μL 1 M $CaCl_2$
- Add dH_2O up to 1 L and autoclave
- 20 mL 100 x BME vitamins (Sigma-Aldrich)
- 0.5 g $^{15}NH_4Cl$ (dilute in 2.5 mL H_2O and use 0.2 μm filter)

Autoinduction media

- 10 g Tryptone
- 5 g Yeast extract
- 3.3 g $(NH_4)_2SO_4$
- 6.8 g KH_2PO_4
- 7.1 g Na_2HPO_4
- 0.5 g Glucose
- 2.0 g α-Lactose
- 0.15 g $MgSO_4$
- Add dH_2O up to 1 L and autoclave

7.1.7 Plasmids and primers

The plasmids for quartromicin M5 (KS-AT-ACP), M3_ACP and M4_ACP were produced by Dr. Orestis Lazos (Challis group) using fosmid libraries previously produced by Malek Zerikly from the following strains:

- *Amycolatopsis orientalis* ATCC 53884 (American Type Culture Collection)[143]
- *Amycolatopsis mediterranei* NRRL 18815 (Agricultural Research Service Patent Culture Collection).[143]

Further constructs were subcloned from M3_ACP, M4_ACP and M5 using primers ordered from IDT (Table 7.3). Primer suitability was checked using OligoAnalyzer 3.1 (IDT). Stocks of primers were diluted to 100 μM (dH_2O) and stored at -20 °C. Small aliquots of

primers were further diluted to 10 μM before use (dH_2O).

Table 7.3. *Primers for all constructs used in this thesis. Bases shown in red correspond to stop codons and bases shown in blue allow the ligation of the PCR product into a TOPO vector. Site-directed mutations are also highlighted.*

Protein	Plasmid name	Origin gene	Length (bp)	Vector	Forward primer	Reverse primer
M3_ACP	pTG5000	*qmnA1*	444	pET151	CACC TTG CTG GAC CTG GTC	CTA TCT TCT CCC GTT GCC
M3_ACPT	pTG5001	*qmnA1*	240	pET151	CACC TTG CTG GAC CTG GTC	CTA GTC GGT GGG TTC CGG
M4_ACP	pTG5002	*qmnA2*	444	pET151	CACC CTG CTC ATC GAC GT	CGG TGC TCA TGC GGT TT
M4_ACPT	pTG5003	*qmnA2*	231	pET151	CACC CTG CTC ATC GAC GTG	CTA CGG GGC CAG TTC GG
M5_ACP	pTG5004	*qmnA3*	261	pET151	CACC ACG GGT GCC GAG CA	CTA GGA TGG TTC GGC GGC
M5_KS	pTG5005	*qmnA3*	1,377	pET151	CACC ATG AGC ACC GGC CCG AC	CTA CCG TTC CAG TAT CAC GTG CGC GTT
M5_KSATS	pTG5006	*qmnA3*	2,664	pET151	TAA TAA CGG GTC GAC CTG CCG ACG	CCT GGC CCC GCT GAA CCC
M5_KSAT	pTG5007	*qmnA3*	3,594	pET151	TAA AAG GGC GAG CTC AGA TCC GG	TTA CGG CAC GCT CGC GCT CAG
M5	pTG5008	*qmnA3*	3,876	pET151	CACC ATG AGC ACC GGC CCG AC	CTA GGA TGG TTC GGC GGC
C207A	pTG5008-11	*qmnA3*	3,876	pET151	CGT CAC CAT CGA CAC CGC AGC CTC GTC CTC	GCC GGT CCC TCC AGG CCG AAG GTG TAG G
S656C	pTG5012-16	*qmnA3*	3,876	pET151	GCC GGT CAT TGC ATC GGC GAG C	ACA AAG GAC GGT CGA ACG C
M5_KSATS-A	pTG5017	*qmnA3*	2,761	pET151	TAA TAA CTC GCC GCC GAA CCA TCC	CAG GTC GAC CTG CGC CGA
M5_KSATS-B	pTG5018	*qmnA3*	3,202	pET151	TAA TAA CTC GCC GCC GAA CCA TCC	CAG GTC GAC CTG CGC CGA
M5_KSATS-C	pTG5019	*qmnA3*	3,382	pET151	TAA TAA CTC GCC GCC GAA CCA TCC	GAG TTC CGC GCG CTC CAG
M5_KSATS-D	pTG5020	*qmnA3*	3,418	pET151	TAA TAA GCC TGG GAC GGC GTC GCT	GAC CAC CAC CGC ACC CCG
M5_KSATS-E	pTG5021	*qmnA3*	3,480	pET151	TAA TAA GCT GCG GCA GGG CGC	ACC AGG CCG ACC GCG CTC G
M5_KSATS-F	pTG5022	*qmnA3*	3,531	pET151	TAA TAA CTC GCC GCC GAA CCA TCC	CAA CGG CGC CAA CGC CTC
M5_KR^D-A	pTG5023	*qmnA3*	717	pET151	CTG GCT GGT CCC GTC CGG	GTG AAG GGA TCA ATT CCC TGA AAA TAC AGG TTT TC
M5_KR^D-B	pTG5024	*qmnA3*	555	pET151	GTA CGC CTG CGA CCA CCT	GTG AAG GGA TCA ATT CCC TGA AAA TAC AGG TTT TC
M5_KR^D-C	pTG5025	*qmnA3*	408	pET151	CGA CGG CGA CTG GTG GCC	GTG AAG GGA TCA ATT CCC TGA AAA TAC AGG TTT TC
M5_KR^D-D	pTG5026	*qmnA3*	288	pET151	ACT GGC GGC GGT TGT CCA	GTG AAG GGA TCA ATT CCC TGA AAA TAC AGG TTT TC

A Champion^{TM} pET151 Directional TOPO® host-expression vector (pET151, Figure 7.1)[191] was used for the overproduction of all proteins. Plasmids in pET151 undergo low-copy replication due to the presence of the basis of mobility (bom) region from pBR322_*origin* (origin of replication) and repressor of primer (*ROP*) genes. The presence of a β-lactamase gene (AmpR) ensures the production of an enzyme that inactivates β-lactam antibiotics, such as ampicillin, thus conferring an antibiotic resistance to *E. coli* cells containing the plasmid. This allows selection through lysogeny broth (LB) agar ampicillin plates of plasmid-containing cells, minimising the chance of selecting an empty vector.

The vector also contains genes encoding a T7 promoter site to which T7 RNA polymerase

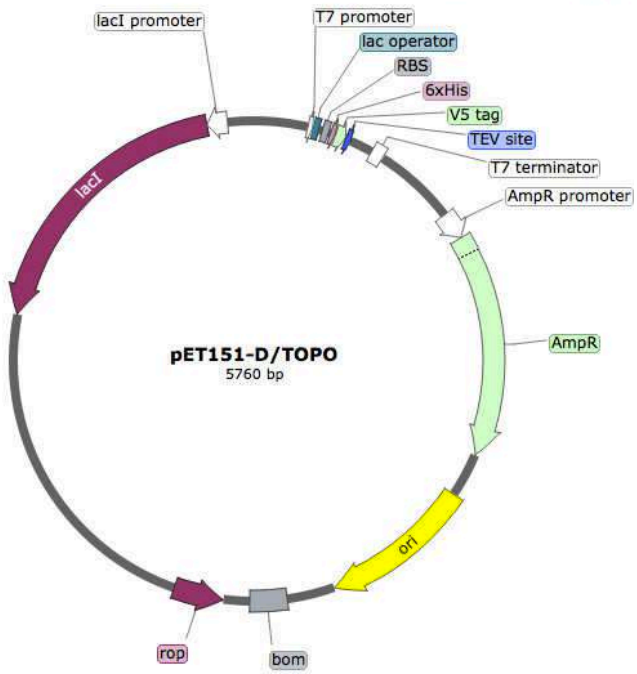


Figure 7.1. *A gene map of the pET151 D/TOPO vector (SnapGene).*

must bind to allow transcription of the GOI to occur. This transcription is regulated by the *lactose* (*lac*) operon, which encodes a repressor of the *lac* operator (*lacI*), preventing the binding of the T7 RNA polymerase to the T7 promoter site. As a result, overexpression of the gene cannot occur whilst the repressor is present, preventing leaky expression of the gene.

In addition, the presence of 6xHis encodes an in-frame His-tag at the protein's N-terminus. A TEV site (ENLYFQG) is also encoded to allow the cleavage of a His-tag from the purified recombinant protein, leaving an N-terminal tag of only 6 residues attached to the protein. A gene encoding a V5 epitope tag on the protein's N-terminus is also present (V5 tag), allowing the tagged protein to be analysed by immunochemical methods, such as Western blot.

pJET1.2 also used to allow successful site-directed mutagenesis of longer constructs due to its significantly shorter backbone. The inclusion of restriction sites allowed the mutated inserts to be transferred back into the original pET151 vector.

7.2 Molecular Biology

7.2.1 Preparation of chemically competent *E. coli*

All proteins were recombinantly overproduced using either One Shot® BL21 Star™ (DE3), NiCo21 (DE3), OverExpress C43 (DE3) or Rosetta™ 2 (DE3) Singles™ strains of *E. coli* (Table 7.2). Plasmids were produced in the One Shot® TOP 10™ strain. All strains of *E. coli* were produced in the lab from commercially available stocks and all solutions, Eppendorfs® and pipette tips were sterile.

10 μL of frozen *E. coli* glycerol stocks were used to inoculate LB (10 mL) and grown overnight (37 °C, 180 rpm). 1 mL of the overnight culture was then used to inoculate LB (100 mL) and grown (37°C, 180 rpm) to $OD_{600} \sim 0.4$ (3 - 4 hours). Cells were recovered by centrifugation (3000 rpm, 15 minutes, 4 °C) and the supernatant decanted. The pellet was resuspended in 10 mL of ice-cold wash buffer (0.1 M $MgCl_2$) and left on ice (20 minutes). The cells were then centrifuged (3,000 rpm, 15 minutes, 4 °C) and the supernatant was decanted, before the pellet was resuspended in 2 mL of ice-cold buffer (0.1 M $CaCl_2$). After being left on ice (20 minutes), glycerol was added (final concentration of 10 %), the cells were aliquoted (50 μL) and flash frozen (liquid nitrogen). Cells were immediately stored at -80 °C.

7.2.2 Transformation of chemically competent *E. coli*

1 μL of plasmid DNA was added to 50 μL of chemically competent cells. The suspension was gently mixed and left on ice (30 minutes). The suspension was heat shocked (1 minute at 42 °C, then 2 minutes on ice), before the addition of 950 μL of LB and incubation (37 °C, 180 rpm, 1 hour). Transformants were then plated on LB agar plates with the relevant antibiotic resistance and incubated overnight (37 °C).

7.2.3 Plasmid isolation

The following method was used to prepare all plasmids used in protein overproduction.

After transforming plasmids into TOP10 *E. coli* and growing on antibiotic plates a resulting colony was picked and used to inoculate 10 mL of LB. The culture was incubated overnight (37 °C, 180 rpm) and harvested by centrifugation (4,000 rpm, 15 minutes). Plasmid isolation was then carried out using a GeneJET Plasmid Miniprep Kit #K0502, #KO503 (Thermo Scientific) according to the manufacturer's instructions, with the exception that dH_2O was used to elute the DNA in place of the provided elution buffer. The concentration of the purified DNA solution was determined by a NanoDrop Lite spectrophotometer (Thermo Scientific) and stored at -20°C.

The pellet was resuspended in 250 μL of resuspension solution and transferred to a microcentrifuge tube where 250 μL of lysis solution was added. 350 μL of neutralisation solution was then added and mixed, before centrifugation (14,000 rpm, 5 minutes). The supernatant was transferred to a GeneJET spin column and centrifuged (14,000 rpm, 1 minute). 500 μL of wash solution was added, before again centrifuging (14,000 rpm, 1 minute). The wash and centrifuge steps were repeated and the GeneJET spin column was transferred into a fresh Eppendorf® tube. 50 μL of dH_{20} was added and the solution left for 2 minutes, before being centrifuged (14,000 rpm, 2 minutes). The flow-through was discarded and the concentration of the purified DNA solution was determined by a NanoDrop Lite spectrophotometer (Thermo Scientific). The purified DNA solution was stored at -20 °C.

7.2.4 Polymerase chain reaction

All polymerase chain reactions (PCR) were done using the Mastercycler® nexus X2 (Eppendorf). Q5 polymerase (New England Biolabs) was used to ensure high-fidelity copying. Each reaction contained between 4 and 8 % DMSO due to the high GC-content of the template DNA. Other reagents and conditions are listed in Table 7.4. A temperature gradient was used to find the most suitable annealing temperature (10 μL of PCR mix at 5 different temperatures) before 50 μL of PCR mix was run with the annealing temperature that produced the optimal results.

Table 7.4. *PCR components and conditions.*

PCR components	Amount (μL)	Final concentration
Q5 reaction buffer (5x)	4	1x
Q5 high-GC enhancer (5x)	4	1x
Q5 polymerase (2000 u/mL)	0.5	1 unit
Forward primer (10 μM)	2.5	500 μM
Reverse primer (10 μM)	2.5	500 μM
Template DNA (100 ng)	0.5	2.5 ng/μL
Dimethyl sulfoxide (DMSO)	4 - 8 %	58 - 68
Total volume	20	

Step	Temperature (°C)	Time
1. Initial denaturing	98	2 minutes
2. Denaturing	98	15 seconds
3. Annealing	55-68	30 seconds
4. Extension	72	1 minute per kbp
Return to step 2 (30 cycles)		
5. Final extension	72	10 minutes

7.2.5 Site-directed mutagenesis

Mutants and a number of truncated constructs were created using the Q5 site-directed mutagenesis kit (New England Biolabs). Primers were designed using NEBaseChanger (New England Biolabs) and PCR was carried out as per manufacturer's guidelines, with extended extension steps set to 8 minutes. PCR products were incubated with a kinase-ligase-Dpn1 (KLD) enzyme mix (10 minutes, room temperature, followed by 20 minutes at 37 °C) to allow complete digestion of the template plasmid by Dpn1 (Table 7.5).

7.2.6 Analysis of PCR products

To analyse the PCR products either 10 or 50 μL of sample was mixed with 2 or 10 μL of 6x DNA loading dye, respectively. The samples were then loaded onto a 1 % agarose gel with a ladder (GenerRuler^{TM} 1kb or high range, Thermo Fisher) and run in a BioRad tank containing 1x TBE buffer, powered by a Power PAC 300 (80V, 1 hr). Gels were visualised and photographed using a UVP imaging system UV transilluminator.

Table 7.5. *PCR components and conditions used to produce mutants and truncated constructs using the NEB-SDM kit.*

Components	Amount (μL)
2x Q5 HF master mix	10
Forward primer (10 μM)	1
Reverse primer (10 μM)	1
Template DNA	1
Nuclease free H_2O	7
Total Volume	20

Step	Temperature (°C)	Time
1. Initial denuturing	98	2 minutes
2. Denaturing	98	15 seconds
3. Annealing	55 - 72	30 seconds
4. Extension	72	1 minute per kbp
Return to step 2 (30 cycles)		
5. Final extension	72	10 minutes

Components	Amount (μL)
PCR product	2
KLD reaction buffer (10x)	1
KLD enzyme mix	1
NF−H_2O	6
Total Volume	10

7.2.7 Gel extraction of PCR products

DNA of the appropriate size was excised from the agarose gel using a GeneJet™ Gel Extraction Kit #K0691 #K0692 (Thermo Scientific). All buffers were provided in the kit and centrifugation was done at 11,000 rpm. The gel slice containing the DNA was cut out and incubated (50 - 60 °C, 10 minutes) with a 1:1 volume of binding buffer. The gel was then vortexed. 800 μL of solubilised gel was transferred into a GeneJet column and centrifuged (1 minute). The flow-through was decanted and 100 μL of binding buffer was added, followed by centrifugation (1 minute) and decantation of the flow-through. 700 μL of wash buffer was added, followed by centrifugation (2 minutes), before the GeneJet column was transferred to a fresh Eppendorf® and left to stand in 50 μL of 60 °C dH_2O (5 minutes) before a final centrifugation step (2 minutes) was carried out. The concentration of the extracted PCR product was determined by a NanoDrop Lite spectrophotometer (Thermo Scientific) and stored at −20 °C.

7.2.8 Ligation of amplicons into a pET151 vector

The gel extracted PCR product was ligated into a pET151 vector using a Champion^{TM} pET Directional TOPO^{TM} Expression Kit (Table 7.6). Reactions were incubated (20 minutes, room temperature) and transformed, plated and grown overnight in LB with 100 μg/mL ampicillin.

Table 7.6. *Components of the ligation reaction between the PCR product and the pET151 vector.*

Component	Volume (μL)
Purified PCR product	0.5 - 4
Salt solution	1
dH_2O	up to 5
pET151 vector	1
Final volume	6

7.2.9 Blunt-end ligation of amplicons into pJET1.2

Gel extracted PCR products were blunt-end ligated into pJET1.2 using the CloneJET PCR Cloning Kit (Thermo Scientific). The reaction mix (Table 7.7) was vortexed (5 seconds) and incubated (22 $^\circ$ C, 30 minutes). 5 μL of ligated product was transformed into TOP10 (as previously described), selected by ampicillin plates, grown and miniprepped.

7.2.10 Restriction digests

8 μL of mini-prepped DNA was digested with 1 μL of the relevant restriction enzyme and 1 μL of the corresponding buffer. Double digests were either done sequentially or through a double digest where an appropriate buffer was available.

Table 7.7. *Components of the blunt-end ligation reaction between the PCR product and the pJET1.2 vector.*

Component	Volume (μL)
Purified PCR product	1
2 x Reaction Buffer	10
dH_2O	up to 17
DNA blunting enzyme	1
pJET1.2 blunt cloning vector	1
T4 DNA ligase	1
Final volume	20

Samples were incubated (37 °C, 1 hr) before 2 μL of 6x loading dye was added. The samples were then loaded onto a 1 % agarose gel with a high range ladder and run in a BioRad tank, powered by a Power PAC 300 (80V, 1 hr). Gels were visualised and photographed using a UVP imaging system UV transilluminator. Extraction of digested DNA was carried out as previously described (7.2.7).

7.2.11 Sequencing

All sequencing was done through GATC-biotech's SUPREMERUN (Sanger sequencing) service. The T7 and pET-RP primers were utilised to sequence the start and end of each construct, with further primers designed to sequence the inner regions of the construct where necessary (Table 7.8).

Table 7.8. *Forward sequencing primers for qmnA3.*

Start point of primer (bp)	Primer sequence
760	GCC GAC GGC CGG TGC AAG
1551	GCG CAC CTC GTT CTC CCA C
2344	CCG GTG GTG TCC ACG GTG A
3140	TCC GTG GTC CGG AGG TGT TC

7.3 Overproduction and purification of proteins

The following steps were used to successfully overexpress and purify all proteins used in this thesis. All steps after lysis were either carried out in a cold room (4 °C) or on ice.

7.3.1 Overproduction of recombinant protein in LB media

Plasmids were transformed into an overproduction strain of *E. coli*, plated overnight (37°C) on ampicillin plates and the resulting colonies were used to inoculate 10 mL cultures of LB with ampicillin (100 μg/mL). The culture was grown overnight (37 ° C, 180 rpm) and was used to inoculate 1 L of LB with ampicillin (100 μg/mL) which was grown (37 °C, 180 rpm) to $OD_{600} \sim 0.6$ (3 - 4 hours). IPTG (500 μM) was then added and the culture was incubated overnight (15 °C, 180 rpm).

Autoinduction media was also used as a growth media in place of LB for less soluble proteins. 10 mL overnight cultures were grown in LB (as above) and used to inoculate 1 L of autoinduction media with ampicillin (100 μg/mL). The culture was then grown (37 °C, 180 rpm) to $OD_{600} \sim 1-2$ ($\sim$ 5 hours), before the temperature was dropped to 20 °C for a further 24 hours of incubation (180 rpm).

Isotope labelling of proteins was essential for facilitating 2D-NMR experiments. To iso-

tope label proteins, 10 mL overnight cultures were grown in LB (as above) and used to inoculate 1 L of LB with ampicillin (100 μg/mL). The culture was then grown (37 °C, 180 rpm) to $OD_{600} \sim 0.6$ (2 - 4 hours) and centrifuged (5000 rpm, 15 minutes, 4 °C). To produce [U-^{15}N] proteins, the pellet was resuspended in 200 mL of M9 salt solution and the cells were centrifuged as before and resuspended in 1 L minimal media (100 μg/mL ampicillin, 0.5 mM IPTG). The culture was then incubated overnight (20 °C, 180 rpm).

7.3.2 Sodium dodecyl sulfate-polyacrylamide gel electrophoresis

Table 7.9. *Recipe for SDS-PAGE gels (6 %, 8 % and 15 %).*

Constituents (mL)	6 % resolving gel	8 % resolving gel	15 % resolving gel	5 % stacking gel
Distilled water	2.6	2.3	1.1	0.68
29:1 (30 %) Acrylamide:Bisacrylamide	1.0	1.3	2.5	0.17
Tris buffer	1.3 (pH 8.8, 1.5 M)	1.3 (pH 8.8, 1.5 M)	1.3 (pH 8.8, 1.5 M)	0.13 (pH 6.8, 1.0 M)
10 % SDS	0.05	0.05	0.05	0.01
10 % Ammonium persulfate	0.05	0.05	0.05	0.01
N,N,N,N-Tetramethylethylenediamine	0.003	0.002	0.002	0.001
Total volume	5	5	5	1

Samples were prepared by mixing 10 μL of protein samples with 2 μL of 6x SDS-PAGE loading dye. 10 μL of sample was then loaded into each lane of the gel. 6 % SDS-PAGE gels were used for the analysis of M5 (141 kDa), 8 % gels for the analysis of the KSAT didomain and 15 % gels for the analysis of ACP and KR^D domains. Constituents were mixed in proportions as shown in Table 7.9, before being set in the gel mould (20 minutes). 70 % ethanol was used to ensure the level of the gel was consistent, before a stacking layer was prepared and added to the top of the gel. The comb was immediately inserted into the gel and it was again left to dry (20 minutes). The gel was then clamped and placed in a Mini-PROTEAN® Tetra Cell System (BioRad) containing 1x protein running buffer. Samples were then loaded into each lane, with PageRuler Plus prestained protein ladder

(Thermo Scientific) as a marker. The gel was run (180 V, 1 hr) and then stained for several hours using InstantBlue Protein Stain (Expedeon). Gels were then destained overnight in H_2O with 1 % ethanol.

7.3.3 Purification of His-tagged proteins by immobilised metal affinity chromatography

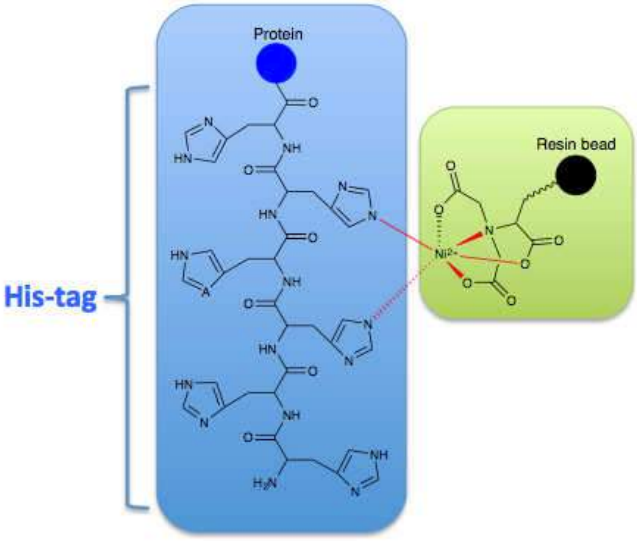


Figure 7.2. *The His-tagged protein binds to the Ni-NTA Sepharose of the HisTrap column.*

The presence in the pET151 vector of a gene encoding an N-terminal polyhistidine (6x) tag (His-tag) allowed the purification of overproduced proteins by immobilised metal affinity chromatography (IMAC). IMAC is a widely used method of His-tagged protein purification. It utilises columns coated with chelating agents that are able to bind and immobilise metals such as nickel, copper and cobalt, through coordinate covalent bonds. The His-tag's consecutive histidine residues have a strong affinity to the immobilised metal's ions, resulting in coordinate bonds forming and the immobilisation of the protein on a nitrilotriacetic acid (NTA) Sepharose column (Figure 7.2). The positioning of the His-tag at the N-termini minimises the possibility of interference in the protein's functionality.

Proteins that had bound non-specifically to the column were removed by washing the column with a buffer containing a low concentration of imidazole, a strong metal ion ligating compound. Imidazole is a compound that is structurally similar to histidine and as a result, is able to form complexes with Nickel (II) ions (Ni(II)). Buffers containing in-

creasing concentrations of imidazole were then passed through the column. Once the imidazole concentration is high enough, stable coordinate bonds are formed between the metal ion and imidazole, resulting in the breaking of the bond between the His-tag and Ni(II). The eluted protein was then collected in fractions.

Overnight cultures of media containing overproduced proteins were centrifuged (11,000 rpm, 20 minutes) and the pellets resuspended in 10 mL of wash buffer. The cells were then lysed using a cell disruptor (Constant Systems E1061), centrifuged (11,000 rpm, 20 minutes) and filtered using a MF-Millipore^{TM} membrane filters (0.2 μm, Merck Millipore) syringe filter.

The cells were purified by using a 1 mL or 5 mL HisTrap™ HP affinity column (Nickel Sepharose High Performance, GE Healthcare Life Sciences). A 5 mL HiTrap chelating HP (GE Healthcare Life Sciences) containing cobalt was used for some M3_ACP samples to improve sample purity. Columns were regenerated before each use (5 mL of stripping buffer, 10 mL of wash buffer and 5 mL of 0.1 M $NiSO_4$ or $CoSO_4$, followed by 20 mL of wash buffer) using a 5 mL syringe. An ÄKTA Pure 25 L (GE Healthcare Life Sciences) fast liquid protein chromatography (FPLC) system was used to pass the sample through the column and the unbound sample was washed off using wash buffer (20 mL). The sample was eluted by a linear gradient with increasing imidazole content (20 mM to 300 mM) and collected in 1.5 mL fractions which were visualised by the ÄKTA's inbuilt UV-system. The length of the gradient varied between 20 mL and 50 mL, depending on whether other proteins coexpressed with a similar affinity to the chelating metal.

A second method of purifying the protein through IMACs was also utilised, with a syringe being used to inject sample to a 1 mL HisTrap column (as used above). The unbound sample was then washed out by wash buffer and buffers containing 50 mM, 100 mM, 200 mM and 300 mM (5 mL of each) were run through the column, in order of increasing concentration, and collected separately. Each fraction was then checked for protein purity by an SDS-PAGE gel.

In both cases, the fractions containing the protein were then concentrated by centrifugation (4,000 rpm, 1 hour) in a Vivaspin® 20 Ultra Centrifugal Filters (5,000, 10,000 or 30,000 MWCO). Glycerol was then added to a final concentration of 10 % before proteins were aliquoted and frozen with liquid nitrogen. Samples were stored at -80 °C.

7.3.4 Prediction of protein concentrations using a Bradford assay

The following assays were run in triplicate. Errors corresponded to one standard deviation.

Proteins were serially diluted to 0.25, 0.5, 0.75, 1.0, 1.25 and 1.5 mg/mL using protein running buffer. BSA was also diluted to the same concentrations. 30 μL of each concentration were then pipetted into separate Eppendorfs, before 1.5 mL of Bradford reagent was added. A blank containing no protein was also prepared. Samples were incubated (15 minutes, room temperature) and absorbance was measured at 595 nm. The BSA results were then used to produce a calibration curve from which the proteins concentration could be estimated.

7.3.5 Cleaving His-tags from proteins

Green fluorescent protein tobacco etch virus (GFP-TEV) protease was used to cleave the His-tag from purified proteins. Assays contained protein and GFP-TEV protease in an approximately 100:1 ratio and were incubated (2 hours, room temperature). TEV digested proteins were run through a nickel-affinity column to remove all remaining His-tagged proteins and the GFP-TEV protease. The flow-through was then collected and examined on an SDS-PAGE gel against the uncleaved protein. Samples were then concentrated and stored (as above).

7.3.6 Tryptic digestion of proteins

Trypsin digests were carried out using the Trypsin Profile IGD Kit For In-Gel Digestions (PP0100, Sigma-Aldrich). All buffers and reagents mentioned are part of this kit. An SDS-PAGE gel was run and the band of interest was cut out using a razor blade. The gel piece was then placed in a prewashed Eppendorf® tube (100 μL of Peptide Extraction Solution)

and covered in 200 μL of Destaining Solution. After incubation (30 minutes, 37 °C), the solution was discarded and this step was repeated. The gel piece was then dried (30 minutes) in a desiccator attached to a HI-VAC® vacuum valve (Sigma-Aldrich).

20 μL of Trypsin Solution (0.4 μg of trypsin) and 50 μL of Trypsin Reaction Buffer was then added to the gel piece, followed by incubation (37 °C, overnight). The solution was then removed from the gel piece, concentrated using a C18 ZipTip® (Merck Millipore, see below) and examined for extracted tryptic peptides by matrix-assisted laser desorption ionisation-time of flight(MALDI-TOF).

7.3.7 Deglycosylation of proteins

Glycan analysis was carried out using the GlycoProfile™ I Enzymatic In-Gel N-Deglycosylation Kit (Sigma-Aldrich) and all buffers used are from this kit.

An SDS-PAGE gel was run and the band of interest was cut out using a razor blade. The gel piece was then washed and dried (7.3.6). 10 μL of PNGase F solution was then added to the gel piece, followed by a brief centrifugation and incubation (37 °C, 30 minutes). 20 μL of dH_2O was added to the gel piece, which was then incubated (37 °C, overnight). The sample was then briefly centrifuged and the solution was removed and retained for glycan analysis (MALDI).

The gel piece was incubated (37 °C, 30 minutes) in 200 μL of dH_2O to remove excess PNGase. The solution was then discarded and the previous step repeated. The deglycosylated protein may then either be extracted from the gel piece for analysis or digested using the above trypsin protocol. The intact deglycosylated protein was extracted from the gel using the following method. The gel piece was dried in a desiccator attached to a vacuum pump (HI-VAC, 45 minutes). 50 μL of Peptide Extraction Solution was added and the sample was incubated (37 °C, 30 minutes). The supernatant was separated and concentrated using the ZipTip® protocol (7.3.8). Samples were analysed by MALDI.

7.3.8 Concentrating proteins and peptides using ZipTip®

The ZipTip was wet with 10 μL of the elution solution (50 % acetonitrile (ACN): 50 % dH_2O), was then dispensed to waste (repeated several times). Equilibrated was carried out by washing three times with 10 μL of wash solution 1 (0.1 M triethylammonium acetate (TEAA), pH 7.0). The sample was bound to the ZipTip by aspirating the sample 10 times, before being washed several times with 10 μL of wash solution 1 followed by wash solution 2 (dH_2O). The wash solution was discarded after each wash. 2 μL of the elution solution was then aspirated and dispensed into a fresh Eppendorf®. Aspiration of the eluted solution was repeated several times to obtain an adequate volume of sample for further analysis (MALDI).

7.3.9 Size exclusion chromatography

Gel filtration columns contain a porous matrix of differing sizes to which the proteins have varying access depending on their size and steric conformation (Figure 7.3). By passing protein samples through a calibrated column and monitoring the elutants using a UV spectrophotometer, it is possible to collect fractions containing the desired purified product. This removes the majority of proteins of incorrect size and any unfolded proteins, as well as confirming the oligomerisation states of the eluted protein.

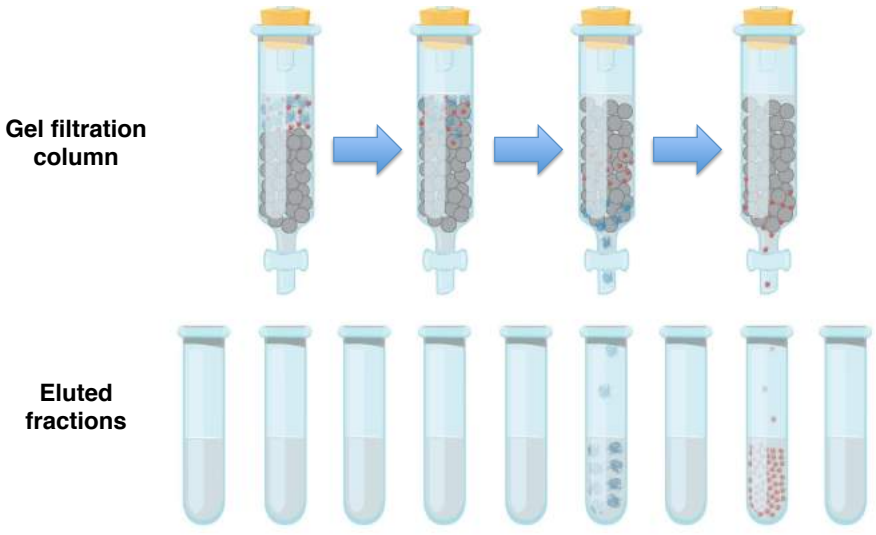


Figure 7.3. *Size exclusion chromatography utilises a porous matrix to separate proteins based upon their size and steric conformation. The protein sample passes through the column at a different rate, depending on the number of pores entered, resulting in larger proteins being eluted before smaller proteins.*

A Superdex^{TM} 200 10/300 GL gel filtration column (GE Healthcare Life Sciences) was washed with 150 mL of storage buffer at 0.5 mL/min on an ÄKTA Pure 25 L (GE Healthcare Life Sciences) FPLC system. The column was then calibrated using protein markers (Table 7.10, 2 mg of each marker, 150 mL of wash buffer), before up to 1 mL of protein was injected. The column was then washed with a further 150 mL storage buffer and the eluted volume was collected in 5 mL fractions. The chromatogram was recorded and used to determine which fraction contained the eluted protein and the corresponding fractions were concentrated and stored (as above). Calibration of a Superdex^{TM} 200 10/300 GL gel filtration column (GE Healthcare Life Sciences) was carried out in the same way with the Gel Filtration Markers Kit for Protein Molecular Weights 6,500 - 66,000 Da (Sigma Aldrich). Elution times of each marker are displayed in Table 7.11.

Table 7.10. *Protein markers used to calibrate the GE Healthcare Life Sciences Superdex^{TM} 200 10/300 GL gel filtration column.*

Protein marker	Molecular weight (kDa)	Elution volume (mL)
Thyroglobulin	660	46
Apoferritin	481	50
β-amylase	200	64
Bovine serum albumin (BSA)	66	76
Lysozyme	14.3	109

Table 7.11. *Protein markers used to calibrate the GE Healthcare Life Sciences Superdex^{TM} 75 10/300 GL gel filtration column.*

Protein marker	Molecular weight (kDa)	Elution volume (mL)
Blue Dextran	2,000	45.2
Bovine serum albumin (BSA)	66	54.3
Carbonic Anhydrase	29	68.5
Cytochrome C	12.4	80.6
Aprotinin	6.5	96.5

7.3.10 Anion exchange chromatography

A HiTrap Q HP (Q Sepharose®, GE Healthcare Life Sciences) was used to further purify proteins. A 5 mL syringe was used to pass samples through the column, before 10 mL of storage buffer was used to wash off the sample not bound to the column. AEX buffers were then passed through the column (5 mL) with increasing NaCl concentrations. Each fraction was collected separately and the location of the protein was determined by SDS-PAGE gel electrophoresis, before the corresponding fractions were concentrated and stored (as above).

7.3.11 Buffer exchange and lyophilisation

A PD10 desalting column (GE Healthcare Life Sciences) was used to transfer protein to a new buffer. The column was washed with 10 mL of the new buffer before up to 2.5 mL of sample was embedded in the column. 10 mL of the replacement buffer was then washed through the column and 0.5 mL fractions were collected. The absorbance of the fractions were then measured using a NanoDrop Lite Spectrophotometer (Thermo Scientific) and peaks were run on an SDS-PAGE gel to confirm which contained the protein. Fractions containing the proteins were then concentrated using a Vivaspin® 20 Ultra Centrifugal Filter column (3, 5, 10, 30, 50 or 100 kDa MWCO, GE Healthcare Life Sciences).

Samples requiring lyophilisation were first transferred into dH_2O. Fractions were concentrated (as above) to a volume of less than 1 mL and transferred to a 1.5 mL Eppendorf tube. The lid was opened and a layer of parafilm placed over the opening. The parafilm was then pierced several times and the sample was flash frozen in liquid nitrogen. All flasks connected to the VirTis BechTop™ K were then isolated and a 600 mL VirTis™ flask containing the Eppendorf was connected to a free valve. All flasks were then reconnected to the instrument and the sample was left overnight to freeze-dry.

7.3.12 Homology modelling of domains, structural features and protein complexes

All homology models in this study were constructed using the Phyre2 server. The utilised templates were chosen by Phyre2's algorithms.

7.4 Analytical techniques

7.4.1 MALDI-TOF

MALDI-TOF was used to confirm the masses of proteins and peptides. A saturated sinapinic acid solution was prepared in 50 % acetonitrile and mixed (1:1) with the sample. A target on a plate (Bruker MTP 384 target plate ground steel T F) was spotted with 0.5 μL of the solution. Samples were then examined using an Autoflex speed MALDI-TOF/TOF spectrometer (Bruker, see manual for settings).

7.4.2 Identifying the location of posttranslational modifications using peptide mass fingerprinting

Intact proteins were trypsin digested (7.3.6) and the resulting peptides were examined using MALDI-TOF (7.4.1). The masses of both the digested protein and the digested protein with PTMs were then compared with the theoretical masses of the enzymatically digested protein (ExPaSy PeptideCutter) to confirm the location(s) of PTMs.

7.4.3 Liquid Chromatography Quadrupole Time Of Flight (LC-QTOF) of intact proteins

All mass spectrometry measurements described in this study were performed using a maXis II superscriptTM (Bruker Daltronics) or an amaZon X (Bruker Daltronics) LC/MS.

Both instruments utilise electrospray ionisation (ESI) to produce ions by applying a high voltage to a liquid to produce an aerosol.

Samples were diluted to 4 μM with sterile water and 100 μL was placed inside an LC/MS vial (Agilent). The protocol was run with the following conditions:

Table 7.12. *LC/MS running conditions.*

Time (minutes)	Buffer A	Buffer B
0	95 %	5 %
10	95 %	5 %
30	0 %	100 %
40	0 %	100 %
40	95 %	5 %
50	95 %	5 %

Parameter	MaXis II	AmaZon X
Column type	C4	C3
Prerun (minutes)	15	5
Injection volume (μL)	2	50
Flow rate (mL/min)	0.05	1.0
LC wavelength (nm)	190 - 400	254
MS range (m/z, Da)	195 - 3,000	100 - 3,000

The MaXis II^{TM} (Bruker) was used with an ACE 3 C4-300 column (2.1 x 100 mm) and the amaZon X with a ZORBAX 300SB-C3 column (4.6 x 150 mm). A blank sample containing LC/MS buffer A was run before and after each set of samples.

The LC chromatogram was then used to identify the retention time of the proteins. The relevant peak was selected and the corresponding MS spectra were deconvoluted using Compass DataAnalysis 4.1's (Bruker) maximum entropy deconvolution algorithm.

7.4.4 LC-QTOF of ejected ions

Samples were diluted to 10 μM with sterile water and 100 μL was placed inside an LC/MS vial (Agilent) before 1 μL was injected into the maXis II^{TM} (Bruker). Running conditions

were as above (Table 7.4.3). Compass DataAnalysis 4.1 (Bruker) was to produce an extracted ion chromatogram (EIC) for the expected mass of the ejected ion ($\pm$ 0.05 Da).

7.4.5 Circular dichroism

Circular dichroism (CD) was used to examine the secondary structure of proteins and to determine at which temperatures and pH they were stable. A 450 μL sample was prepared (approximately 0.1 mg/mL of protein) by diluting in CD buffer. Samples were then placed inside a 1 mm cuvette. The spectra were taken from 260 - 190 nm using the JASCO-1500 spectrometer with a Peltier type cell holder (PTC-423L) to allow for variable-temperature measurements. D.I.T was set to 1 second and the bandwidth as 1.0 nm, with a data pitch of 0.2 nm and a scanning speed of 100 nm/min. For each sample, 8 acquisitions were taken. The spectra were then accumulated, with a reference sample containing only buffer subtracted for calibration.

7.4.6 Variable-temperature CD

Samples were prepared (as above) and a small amount of oil was added to prevent evaporation upon heating. Four spectra were taken for each temperature in 10 °C increments from 10 °C to 70 °C and back down to 10 °C. The spectra for each temperature were then accumulated, with a reference sample containing only buffer subtracted for calibration.

7.4.7 Solution NMR spectroscopy

Nuclear Magnetic Resonance (NMR) was independently developed by Felix Bloch and Edward Mills Purcell,[192;193] resulting in their sharing of the 1952 Nobel Prize in Physics. It uses the fact that all nuclei have a charge and most have an intrinsic property, known as spin, that causes them to behave as a magnet and therefore align under a static magnetic field. The spin of each proton and neutron within an atomic nucleus contribute to its overall spin. If the spins are paired against each other, such as in ^{12}C, then the atomic nucleus does not possess an overall spin. However, if the spins are unpaired, such as in ^{1}H, ^{13}C and ^{15}N, then the atomic nucleus possesses an overall spin and can produce a

dipole moment ($\vec{\mu}$):

$$\vec{\mu} = \sum_i q_i \vec{r}_i,$$

where q_i is the magnitude of the ith charge and $\vec{r}_i$ is the position of the ith charge.

If a magnetic field, B_0, is applied to a sample containing isotopes of odd nucleon mass (spin = $\frac{1}{2}$) then each nucleus aligns with or against the magnetic field. The ratio of nuclei aligning with the field is described by the Boltzmann distribution:

$$p_i = \frac{\exp(-\epsilon_i/kT)}{\sum_{j=1}^{M} \exp(-\epsilon_j/kT)},$$

where p_i is the probability of the system being in state i, T is the temperature, M is the number of accessible states, k is the Boltzmann constant and ϵ_i is the energy state of i.[194]

The direction of B_0 is designated as z and no NMR signal is generated whilst B_0 is in equilibrium along the z-direction. To generate an NMR signal, the equilibrium must be perturbed by the application of a second magnetic field in the form of a radio frequency (RF) pulse. The frequency of the RF pulse is known as the Larmor frequency and the direction of the pulse is designated as the x-direction. This causes the magnetisation to precess around the z-direction and move towards the y-direction (Figure 7.4). The angle at which the magnetic field perturbs is dependent on both the strength and length of the RF pulse.

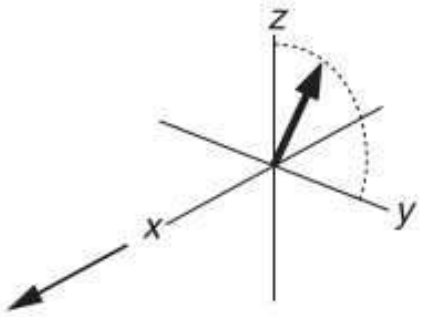


Figure 7.4. *A magnetic field may be perturbed by an RF pulse.*

Upon termination of an RF pulse, the magnetic moment relaxes towards the z-direction. This results in NMR signals becoming weaker and broader over time. The NMR signal intensity may be plotted over time (free induction delay, FID), resulting in the signals being displayed by overlapping sine curves. This may be resolved by Fourier transformation of the FID, resulting in a spectrum in which each peak represents a Larmor frequency of a single nucleus. Modern-day NMR programmes use sequences of pulses to manipulate the nuclei and elucidate the required information.[195]

The RF pulse also causes an energy transition from the base energy to a higher energy level. The spin then returns to the base level, emitting energy at the resonant frequency (v), which is dependent on the effective magnetic field at the nucleus. This can be affected by proximal electrons shielding the nucleus from magnetism. The effect of the shielding leads to a perturbation in the frequency. This is measured against the Larmor frequency (v_0) to give the chemical shift (δ):

$$\delta = \frac{(v - v_0)}{v_0} \cdot 10^6.$$

The concentrated proteins were prepared to 100 - 500 μM in 50 mM phosphate buffer pH 6.5, 5 % D_2O and 0.6 μL of 0.1 M DSS. 600 μL of the solution was pipetted into an NMR tube. An Avance II-700 NMR (AV TwinBay 5 Channel 700) spectrometer was used to acquire all spectra with a 16.45 Tesla ultra-shielded superconducting magnet (700 MHz) and triple resonance cryoprobe (Bruker). The proton resonance frequency was 700.24 MHz and the operating temperature for each experiment was 298 K, unless otherwise stated.

Samples were locked against the deuterium signal, before shimming, matching and tuning were performed. Spectra were acquired by taking 32 scans for each condition, unless otherwise stated. All results were processed with Bruker TopSpin™ and NMRFAM-SPARKY.[196]

7.4.8 Negative-stain electron microscopy

A 5 μL sample of 25 mg/mL protein was placed on a clean microscopic slide, before 5 μL of uranyl acetate was applied. The plate was left to air dry (1 minute) before images were acquired using a Jeol 2010F transmission electron microscope (Ian Hands Portman). Results were analysed using ImageJ version 1.48.[197;198]

Further negative-stain images were obtained at the Beckmann Lab, Genzentrum, LMU (O. Berninghausen) using a 100kV Morgagni microscope, equipped with a side-mounted SIS Megaview 1K CCD camera. Samples were prepared and analysed (as above).

7.4.9 Cryogenic-electron microscopy

Samples were then vitrified using a Vitrobot™ and cryo-EM data was acquired using a 120 kV SPIRIT transmission electron microscope (Tecnai) with an Eagle 2K CCD camera (pixel size of 30 μm).

7.4.10 Surface plasmon resonance

Biacore T2000 surface plasmon resonance (SPR) and Biacore T200 instruments were used to immobilise ligand and collect kinetic data of their interactions with analytes. Ligand proteins were diluted in 10 mM sodium acetate (pH 5.5) and immobilised on a CM5 chip (Biacore T2000) or an S-series CM5 chip (Biacore T200) by amine coupling using 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC), N-hydroxysuccinimide (NHS) and ethanolamine (Figure 7.5). A blank lane was also deactivated, allowing for subtraction of a baseline.

Analyte proteins were then serially diluted in HEPES buffered saline (HBS) and a range of concentrations were run over the CM5 chip lanes with an immobilised ligand and a blank. Data was observed and edited using the Biacore evaluation software. The software was also used to create mathematical kinetic fits to acquired data, with a 1:1 binding model chosen. All kinetics were calculated from a range of concentrations (1.25 - 20 μM),

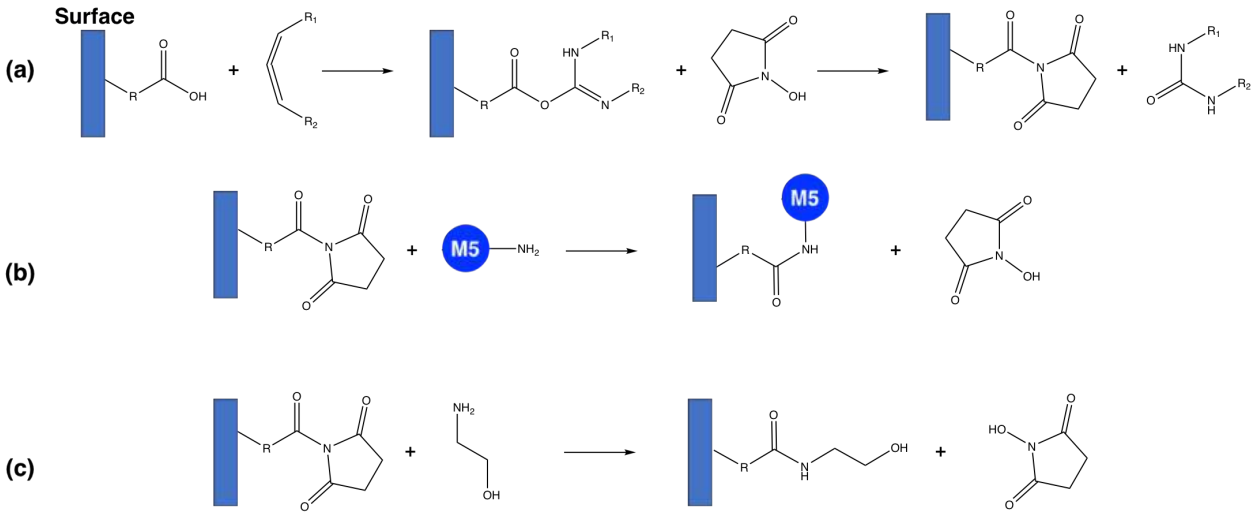


Figure 7.5. *Ligand immobilisation to the surface of a chip. EDC/NHS activates the carboxyl group* ***(a)*** *before the ligand is covalently attached to the surface* ***(b)****. Ethanolamine then deactivates the remaining esters* ***(c)****.* R_1 = $CH_2\,CH_3$, R_2 = $(CH_2)_3\,N^+\,H\,(CH_3)_2\,Cl^-$.

with each sample run in triplicate. Errors given correspond to one standard deviation.

7.5 Biophysical assays

The following assays were all examined using the maXis II^{TM} (Bruker) following the previously described protocols.

7.5.1 Conversion of recombinant ACPs from *apo* to *holo*-forms

Apo-ACPs were posttranslationally modified to *holo*-ACPs using a surfactin phosphopantetheinyl transferase (Sfp, a phosphopantetheinyl transferase (PPTase) from *B. subtilis*), $MgCl_2$ and coenzyme A (CoA-SH). 50 μL aliquots of *apo*-ACPs were pipetted into a 1.5 mL Eppendorf. Samples were then diluted to final concentrations of 200 μM *apo*-ACP, 10 mM $MgCl_2$, 10 μM Sfp, 0.5 mM CoA-SH and 100 mM Tris-HCl (pH 7.5) with storage buffer. The reaction was then incubated (30 minutes, room temperature), after which conversion was confirmed by LC-QTOF (previously described method).

7.5.2 Conversion of recombinant ACPs from *apo* to acyl-*holo*-forms

A range of acyl-substrates were attached to *apo*-ACPs using the following method. The type of acyl group attached to the phosphopantetheine arm was dependent on the type of acyl-CoA used.

50 μL aliquots of *apo*-ACPs were pipetted into a 1.5 mL Eppendorf. Samples were then diluted to final concentrations of 200 μM *apo*-ACP, 10 mM $MgCl_2$, 10 μM Sfp, 0.5 mM acyl-CoA and 100 mM Tris-HCl (pH 7.5). Dilutions were carried out with storage buffer. The reaction was then incubated (2 hours, room temperature), after which conversion was confirmed by LC-QTOF (Table 7.4.3).

7.5.3 Acyl-transfer from M3/M4 _ACP to M5

The reaction mixtures from 7.5.2 were mixed with an equimolar concentration of M5 (unless otherwise stated) and diluted with storage buffer to a final concentration of 50 μM of each protein. The assays were then incubated (overnight, room temperature) and analysed by LC-QTOF (Table 7.4.3). Samples that were collected at time points before acyl-transfer was likely to have completed were immediately frozen at -80° C to prevent further transfer occurring.

7.5.4 Direct competition assay of acyl-transfer from M3_ACP and 4_ACP to M5

Octanoyl/decanoyl-*holo*-M3_ACP and M4_ACP were prepared (as described in 7.5.2). Both reactions were mixed in equimolar amounts and diluted to contain 100 μM of each ACP with storage buffer. M5 was diluted to 100 μM with storage buffer and equal amounts of each sample were then mixed such that the final concentration of each ACP and M5 was 50 μM. The sample was incubated (overnight, room temperature) and analysed by LC-QTOF (Table 7.4.3).

7.5.5 The loading of acyl-substrates onto M5's KS domain and the subsequent extension of a malonyl-derived extender unit

48 μL of the reactant from 7.5.1 was mixed with 1μL of (methyl)malonyl-CoA (20 mM) and 1 μL of acyl-*S*-*N*-acetylcysteamine (SNAC, 20 mM). The extended product depends both on the acyl-SNAC loaded onto the KS of M5 and the malonyl-derived extender unit utilised by the AT of M5. The reactions were then incubated (overnight, room temperature) to test the ability of M5 to catalytically transfer and extend each of these substrates, after which extended products were detected by LC-QTOF (Table 7.4.3).

7.5.6 KS-ACP cross-linking between ACPs and truncated M5 constructs

Cross-linking of M3_ACP, M3_ACPT, M4_ACP, M4_ACPT and M5_ACP with both full length and truncated M5 constructs was achieved by first attaching a ppant cross-linker to the *apo*-ACP before adding M5 in a 2:1 molar ratio (ACP:M5). Each *apo*-ACP was transferred to a 1.5 mL Eppendorf. Attachment of a pantetheine cross-linker to the ACP was then achieved by mixing the ACP with the following constituents to make a 50 μL reaction with the final concentrations as follows: *apo*-ACP (140 μM) with a pantetheine-derived KS-AT cross-linker (0.8 mM), PanK (250 μM), PPAT (250 μM), DPCK (250 μM), $MgCl_2$ (10 μM), Sfp (500 μM) and ATP (2 mM). After incubating (2 hours, room temperature), the increase in mass was confirmed by LC-QTOF.

The cross-linking of the pantetheine-loaded ACP to the KS of M5 was then achieved by adding 50 μL of M5 (140 μL) to the reaction mixture, resulting in a reaction mixture containing the ACP and M5 in a 2:1 molar ratio. The reaction was then incubated (overnight, room temperature) and cross-linking was confirmed using SDS-PAGE gel electrophoresis and LC-QTOF.

7.5.7 AT-ACP cross-linking between ACPs and truncated M5 constructs

Attempts at cross-linking of M5_ACP with M5_KSAT-C207A/S656C, via the AT active site, was achieved using the same method as described in 7.5.6.

7.6 Synthesis of SNAC and pantetheine-bound substrates

The following work was carried out by Matthew Beech, Christian Hobson, Joleen Masschelein and Zhou Shanshan.

7.6.1 Synthesis of protected pantetheines

Pantetheine was synthesised by Matthew Beech using the following method.

5 g of D-pantothenic acid hemicalcium salt was mixed with 1.2 equivalents of p-toluene sulfonic acids (pTSA) in a 1 L round bottom flask. 250 mL of acetone was added and the flask was evacuated with argon and stirred (overnight, room temperature). The resulting slurry was filtered over celite and acetone was removed in a vacuum, yielding a yellow oil. The oil was dissolved in 50 mL of EtOH and washed with brine (3 x 50 mL) before being dried over $MgSO_4$ and concentrated in a vacuum, yielding a white solid. After drying under a high vacuum overnight a yield of 4.029 g was achieved with 89.9 % purity (Figure 7.6).

HRMS: m/z calculated for $C_{12}H_{21}NO_5Na$ $[M^+Na]^+$ = 282.1312 Da, found: 282.1309 Da.

^{1}H NMR (500 MHz, CDCl3) δ**:** 7.05 (bs, 1H, H-10), 4.12 (s, 1H, H-8), 3.71 (d, 1H, ^2J = 12.0, H-4a), 3.66-3.58 (m, 1H, H-11a), 3.56-3.48 (m, 1H, H-11b), 3.31 (d, 1H, ^2J= 12.0, H-4b), 2.65 (t, 1H, ^3J = 6.5, H-12), 1.49 (s, 3H, H-1/2) 1.45 (s, 3H, H-1/2), 1.06 (s, 3H, H-6/7), 1.00 (s,

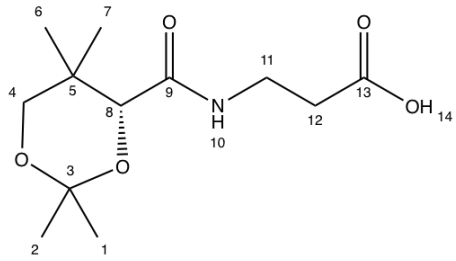


Figure 7.6. *(R)-3-(2,2,5,5-tetramethyl-1,3-dioxane-4-carboxamido)propanoic acid.*

3H, H-6/7).

^{13}C NMR (125 MHz, CDCl3) δ**:** 175.7 (C-13), 170.2 (C-9), 99.3 (C-3), 77.2 (C-8), 71.5 (C-4), 34.1 (C-11), 33.9 (C-12), 33.0 (C-5), 29.4 (C-1/2), 22.0 (C-6/7), 18.9 (C-6/7), 18.7 (C-1/2).

3 g of the resulting product was placed in a 250 mL round bottom flask with 1.1 equivalents of carbonyldiimidazole (CDI). The flask was evacuated with argon and 50 mL of anhydrous tetrahydrofuran was added by cannula. The reaction was stirred for one hour before cysteamine-HCl was added. The flask was then evacuated with argon and stirred (overnight, room temperature). The solvent was then removed by vacuum, the product was redissolved in dichloromethane and washed with saturated NH_4Cl (2 x 50 mL) and brine (2 x 50 mL). The organic phase was dried over $MgSO_4$ and concentrated to give an amorphous solid. The product was then purified by thin layer chromatography (TLC) in 100 % ethyl acetate, yielding 1.701 g of product with 80 % purity (Figure 7.7).

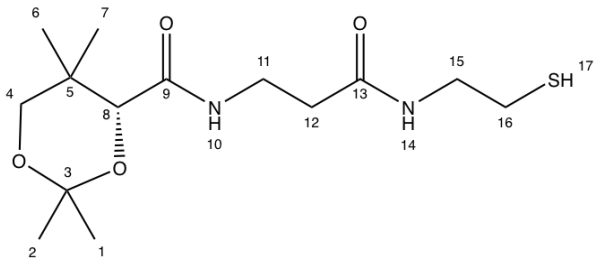


Figure 7.7. *(R)-N-(3-((2-mercaptoethyl)amino)-3-oxopropyl)-2,2,5,5-tetramethyl-1,3-dioxane-4-carboxamide.*

HRMS: m/z calculated for $C_{14}H_{26}N_2O_4SNa$ $[M^+Na]^+$ = 341.1505 Da, found: 341.1505 Da.

^{1}H NMR (500 MHz, CDCl3) δ**:** 7.01 (bs, 1H, H-10), 6.25 (bs, 1H, H-14), 4.09 (s, 1H, H-8), 3.69 (d, 1H, ^{2}J = 12.0, H-4a), 3.63-3.50 (m, 2H, H-11), 3.50-3.36 (m, 2H, H-15), 3.28 (d, 1H, ^{2}J = 12.0, H-4b), 2.69-2.63 (m, 2H, H-16), 2.47 (dt, 2H, ^{3}J = 6.0, 2.0, H-12), 1.46 (s, 3H, H-1/2), 1.42 (s, 3H, H-1/2), 1.37 (t, 1H, ^{3}J = 8.5, H-17), 1.04 (s, 3H, H-6/7), 0.97 (s, 3H,

H-6/7).

^{13}C NMR (125 MHz, CDCl3) δ**:** 171.1 (C-13), 170.4 (C-9), 99.2 (C-3), 77.2 (C-8), 71.5 (C-4), 42.5 (C-15), 36.2 (C-12), 34.9 (C-11), 33.0 (C-5), 29.5 (C-1/2), 24.6 (C-16), 22.2 (C-6/7), 19.0 (C-6/7), 18.7 (C-1/2).

7.6.2 Synthesis of pantetheine bound substrates

The previously synthesised pantetheine (Figure 7.6.1) was coupled to decanoic acid and dodecanoic acid using the following procedure.

50 mg of Figure 7.6.1, 45.2 mg of EDC and substrates (32.4 mg decanoic acid or 37.6 mg dodecanoic acid) were dissolved in dichloromethane (DCM) under argon. 56 μL of N,N-Diisopropylethylamine (DIPEA) was added and the mixture was stirred overnight. After dilution in 5 mL of DCM, the mixture was washed with a 1 M HCl $NaHCO_3$ solution. The solvents were then dissolved under vacuum and dried under HI-VAC. The products were then purified by TLC in 100 % acetate and fractions were dried under HI-VAC to yield 15 mg (decanoic) and 24 mg (dodecanoic) of compound (Figure 7.8).

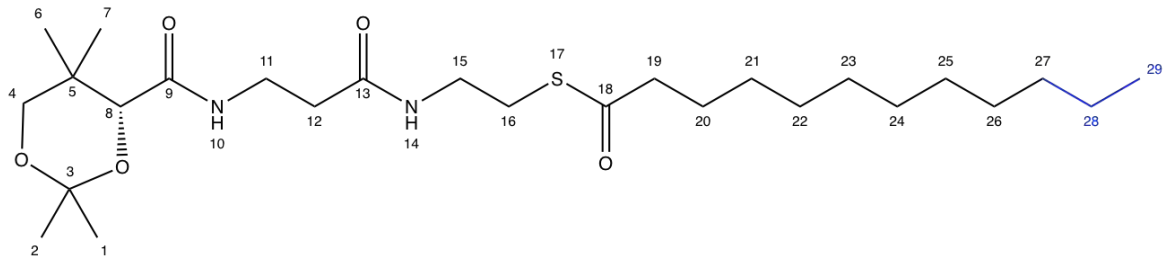


Figure 7.8. *(R)-S-(2-(3-(2,2,5,5-tetramethyl-1,3-dioxane-4-carboxamido)propanamido)ethyl) (do)decanethioate.*

(*R*)-*S*-(2-(3-(2,2,5,5-tetramethyl-1,3-dioxane-4-carboxamido)propanamido)ethyl) decanethioate

HRMS: m/z calculated for $C_{24}H_{45}N_2O_5SNa$ $[M^+Na]^+$ = 495.2863 Da, found: 495.2863 Da.

^{1}H NMR (500 MHz, CDCl3) δ**:** 6.95 (s, 1H, H-10/14), 4.01 (s, 2H, H-15), 4.79 (m, 1H, H-8), 3.59-3.64 (d, 2H, ^2J = 12.5, H-11b), 3.57-3.30 (m, 2H, H-4), 3.24-3.19 (d, 2H, ^2J = 12.0,

H-11a), 2.96-2.92 (t, 2H, ^3J = 6.5, H-16), 2.51-2.47 (t, 2H, ^3J = 12.5, H-12), 2.37-2.34 (t, 2H, ^3J = 6.0, H-19), 1.50 (s, 2H, H-20), 1.20-1.17 (m, H1/2/21-26),0.97 (s, H3, H-6/7/27), 0.90 (s, H3, H-6/7/27), 0.81 (t, H3, ^3J = 7.0, H-6/7/27).

^{13}C NMR (125 MHz, CDCl3) δ**:** 200.0 (C-18), 92.5 (C-8), 71.4 (C-4), 44.2 (C-19), 39.6 (C-12), 36.0 (C-16), 34.8 (C-5), 31.8 (C-25), 29.7-28.5 (C-20 to C-24), 25.6 (C-1/2), 22.7 (C-26), 16.9 (C-6/7).

(*R*)-*S*-(2-(3-(2,2,5,5-tetramethyl-1,3-dioxane-4-carboxamido)propanamido)ethyl) dodecanethioate

HRMS: m/z calculated for $C_{26}H_{48}N_2O_5SNa$ $[M^+Na]^+$ = 523.3175 Da, found: 523.3178 Da.

^{1}H NMR (500 MHz, CDCl3) δ**:** 8.01 (s, 1H, H-10/14), 7.99 (s, 1H, H-10/14), 4.10 (s, 1H, H-8), 3.87 (s, 2H, H-15), 3.71-3.69 (d, 2H, ^2J = 11.5, H-11b), 3.64-3.56 (m, 2H, H-4), 3.32-3.29 (d, 2H, ^2J = 11.5, H-11a), 3.05-3.01 (t, 2H, ^3J = 6.5, H-16), 2.61-2.56 (t, 2H, ^3J = 8.0, H-12), 2.46-2.42 (t, 2H, ^3J = 6.0, H-19), 1.61 (s, 2H, H-20), 1.29-1.27 (m, 2H, H-1/2 and H-21 to H-28), 0.92-0.89 (t, 3H, ^3J = 7.0, H-6/7/29).

^{13}C NMR (125 MHz, CDCl3) δ**:** 199.9 (C-18), 171.2 (C-13), 170.2 (C-9), 114.6 (C-3) 99.0 (C-8), 71.4 (C-4), 44.2 (C-19), 39.6 (C-12), 38.6 (C-11), 36.0 (C-15), 34.8 (C-16), 33.0 (C-5), 31.9 (C-27), 29.8-28.4 (C20-26), 25.6 (C1/2), 22.7 (C28), 18.9 (C6/7), 18.7 (C6/7), 14.1 (C-29).

7.6.3 Deprotection of the SNAC / pantetheine-bound substrates

The protected pantetheines (7.6.2) were diluted in 0.5 mL of tetrahydrofuran (THF) and 10 mL of rubidium fluoride (RBF). 0.5 mL of H_2O and 1 mL of acetic acid were added and the mixture was stirred (72 hours). Solvents were removed under vacuum at 45 °C and the compound was purified by TLC in 5 % methanol in DCM. The removal of solvent under vacuum yielded 6.5 mg (decanoyl pantetheine) and 58 μg (dodecanoyl pantetheine, Figure 8.2).

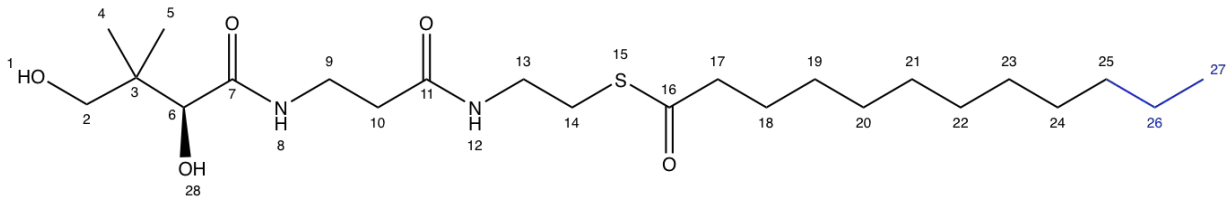


Figure 7.9. *(S)-S-(2-(3-(2,4-dihydroxy-3,3-dimethylbutanamido)propanamido)ethyl) (do)decanethioate.*

(*S*)-*S*-(2-(3-(2,4-dihydroxy-3,3-dimethylbutanamido)propanamido)ethyl) decanethioate

HRMS: m/z calculated for $C_{21}H_{40}N_2O_5SNa$ $[M^+Na]^+$ = 455.2550 Da, found: 455.2553 Da.

^{1}H NMR (500 MHz, CDCl3) δ**:** 7.32 (s, H1, H-8/12), 6.23 (s, 1H, H-7), 5.23 (s, H1, H-1), 3.93 (s, H1, H-3/6), 3.52-3.47 (m, 2H, H-2), 3.42 (s, H1, H-13), 3.35-3.27 (m, 2H, H-9), 3.01-2.89 (m, 2H, H-14), 2.52-2.48 (t, 2H, ^3J = 7.5, H-10), 2.37-2.33 (t, 2H, ^3J = 7.0, H-17), 1.61-1.52 (m, 2H, H-18), 1.26-1.17 (m, 2H, H-19 to H-24), 0.83-0.80 (m, 3H, H-4/5/25).

^{13}C NMR (125 MHz, CDCl3) δ**:** 14.1 (C-27), 20.4 (C-6/7), 21.7 (C-6/7), 22.7 (C-26), 25.6 (C-20), 28.3-29.3 (C-21 to C-24), 31.8 (C-25), 35.2 (C-16), 35.6 (C-15), 39.4 (C-11), 39.7 (C-12), 44.2 (C-19), 70.9 (C-4), 77.3 (C-8), 171.7 (C-13), 173.6 (C-9), 200.5 (C-18).

(*S*)-*S*-(2-(3-(2,4-dihydroxy-3,3-dimethylbutanamido)propanamido)ethyl) dodecanethioate

HRMS: m/z calculated for $C_{23}H_{44}N_2O_5SNa$ $[M^+Na]^+$ = 483.2863 Da, found: 483.2863 Da.

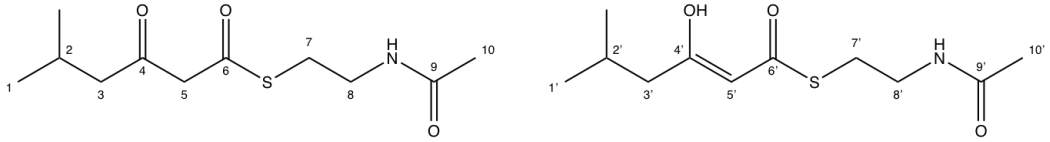
^{1}H NMR (500 MHz, CDCl3) δ**:** 7.37(bs, 1H, H-8), 6.17 (bs, 1H, H-12), 4.02 (s, 1H, H-6), 3.62-3.58 (m, 1H, H-4,H-9), 3.55-3.48 (m, 1H, H-2/13), 3.11-2.99 (m, 2H, H-14), 2.61-2.57 (t, 2H, ^3J = 7.0, H-10), 2.45-2.42 (t, 2H, ^3J = 6.0, H-17), 1.70-1.63 (m, 2H, H-18), 1.28-1.23 (m, 16H, H-19 to H-26), 1.05 (s, 3H, H-4/5), 0.95 (s, 3H, H-4/5), 0.89-0.92 (t, 3H, ^3J = 7.0, H-27).

^{13}C NMR (125 MHz, CDCl3) δ**:** 200.5 (C-18), 173.5 (C-13), 171.7 (C-9), 77.3 (C-8), 71.0 (C-4), 44.2 (C-19), 39.7 (C-12), 35.6 (C-11), 35.1 (C-15), 32.0 (C-17), 28.3-29.6 (C-21 to C-26), 25.7 (C-20), 22.7 (C28), 21.7 (C-6/7), 20.4 (C-6/7), 14.1 (C-29).

7.6.4 Production of SNACs

The following SNAC-bound substrates were provided by various members of the Challis group. All NMR assignments and mass spectrometry data were obtained by the group members who produced the corresponding substrates.

_S_-(2-acetamidoethyl) 5-methyl-3-oxohexanethioate

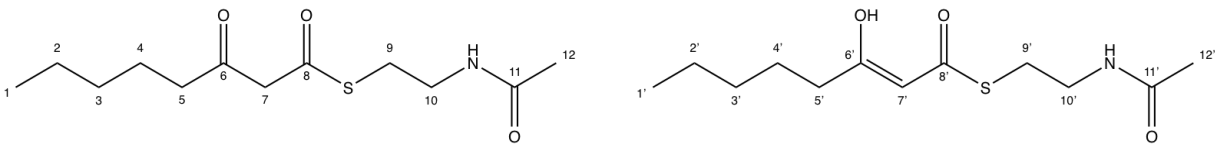


HRMS: m/z calculated for $C_{11}H_{19}NO_3SNa$ $[M^+Na]^+$ = 268.0978 Da, found: 268.0985 Da.

^{1}H NMR (400 MHz, CDCl3) δ: 12.55 (br s, 1H, O-H), 6.23 (br s, 1H, N-H), 5.40 (s, 1H, H-5), 3.65 (s, 2H, H-5′), 3.42 (t, 2H, ^3J = 6.0, H-6), 3.05 (t, 2H, ^3J = 6.0, H-7), 2.37 (d, 2H, ^2J = 7.0, H-3), 2.11 (m, 1H, H-2), 1.99 (br, s, 2H H-3), 1.94 (s, 3H, H-10), 0.89 (d, 6H, ^2J = 6.5, H-1).

^{13}C NMR (100 MHz, CDCl3) δ: 194.08 (C-4′), 176.54 (C-6), 170.51 (C-9), 170.36 (C-9′), 100.04 (C-5′), 57.45 (C-5), 52.10 (C-3), 43.95 (C-3′), 39.74 (C-8′), 39.04 (C-8), 29.04 (C-7), 27.67 (C-7′), 26.33 (C-2), 24.23 (C-10), 22.29 (C-1).

_S_-(2-acetamidoethyl) 3-oxooctanethioate



HRMS: m/z calculated for $C_{12}H_{21}NO_3SNa$ $[M^+Na]^+$ = 282.1134 Da, found: 282.1147 Da.

^{1}H NMR (400 MHz, CDCl3) δ: 12.59 (br s, 1H, O-H), 6.02 (br s, 1H N-H), 5.45 (s, 1H, H-7′), 3.68 (s, 2H, H-7), 3.48-3.42 (m, 2H, H-10), 3.09-3.05 (m, 2H, H-9), 2.51 (t, 2H, ^3J = 7.5, H-5), 2.16 (t, 2H, ^3J = 7.5, H-5′), 1.96 (s, 3H, H-12), 1.58 (q, 2H, ^5J = 7.0, H-4), 1.31-1.24 (m,

4H, H-2/3), 0.90-0.86 (m, 3H, H-1).

^{13}C NMR (100 MHz, CDCl3) δ**:** 194.24 (C-6′), 177.63 (C-8), 170.41 (C-11), 99.09 (C-7′), 57.14 (C-7), 43.37 (C-5), 39.86 (C-5′), 39.14 (C-10), 34.82 (C-10′), 31.19 (C-9′), 31.07 (C-9), 29.17 (C-3), 27.78 (C-3′), 25.86 (C-4′), 23.11 (C-12), 23.05 (C-4), 22.32 (C-2), 22.28 (C-2′), 13.82 (C-1).

Bibliography

[1] S. Marlow, *British Medical Journal*, 2008, **336**, 777.

[2] J. Scarborough, *Pharmacy in History*, 55, **2**, 55–63.

[3] F. W. Serturner and G. Lockemann, *Journal of Chemical Education*, 1951, **28**, 277–279.

[4] J. Miner and A. Hoffhines, *Texas Heart Institute Journal*, 2007, **34**, 179–186.

[5] J. A. Page, *The Aspirin Wars: Money, Medicine, and 100 Years of Rampant Competition. N*, Georgetown University, 1992, pp. 459–481.

[6] A. Fleming, *British Journal of Experimental Pathology*, 1929, **10**, 226–236.

[7] B. Spellberg, J. H. Powers, E. P. Brass, L. G. Miller and J. E. Edwards, *Clinical Infectious Diseases*, 2004, **38**, 1279–1286.

[8] M. Falagas and S. Kasiakou, *Clinical Infectious Diseases*, 2005, **40**, 1333–1341.

[9] P. Acred, D. M. Brown, D. H. Turner and M. J. Wilson, *British Journal of Pharmacology & Chemotherapy*, 1962, **18**, 356–69.

[10] W. Tao, M. Zhu, Z. Deng and Y. Sun, *Science China Chemistry*, 2013, **56**, 1364–1371.

[11] C. Calvori, L. Frontali, L. Leoni and G. Tecce, *Nature*, 1965, **207**, 417–18.

[12] I. S. Johnson, J. G. Armstrong and M. Gorman, *Cancer Research*, 1963, 1390–1427.

[13] G. Bonadonna and S. Monfardini, *The Lancet*, 1969, 1967–1968.

[14] M. Grach, W. Massalha, D. Pud, R. Adler and E. Eisenberg, *British Journal of Clinical Pharmacology*, 2004, **58**, 235–242.

[15] D. G. Brown, T. Lister and T. L. May-Dracka, *Bioorganic & Medicinal Chemistry Letters*, 2014, **24**, 413–418.

[16] B. B. Mishra and V. K. Tiwari, *European Journal of Medicinal Chemistry*, 2011, **46**, 4769–4807.

[17] A. L. Harvey, R. Edrada-Ebel and R. J. Quinn, *Nature Reviews Drug Discovery*, 2015, **14**, 111–129.

[18] J. T. Staley, *Annual Review of Microbiology*, 1985, **39**, 321–46.

[19] A. Penesyan, M. Gillings and I. T. Paulsen, *Molecules*, 2015, **20**, 5286–5298.

[20] World Health Organization, *WHO Press*, 2015, 1–28.

[21] P. McGann, E. Snesrud, R. Maybank, B. Corey, A. C. Ong, R. Clifford, M. Hinkle, T. Whitman, E. Lesho and K. E. Schaecher, *Antimicrobial Agents & Chemotherapy*, 2016, **60**, 4420–21.

[22] J. Masschelein, W. Mattheus, L. J. Gao, P. Moons, R. V. Houdt, B. Uytterhoeven, C. Lamberigts, E. Lescrinier, J. Rozenski, P. Herdewijn, A. Aertsen, C. Michiels and R. Lavigne, *PloS one*, 2013, **8**, e54143.

[23] T. Martens, L. Gram, H. P. Grossart, D. Kessler, R. Müller, M. Simon, S. C. Wenzel and T. Brinkhoff, *Microbial Ecology*, 2007, **54**, 31–42.

[24] E. J. Drake, B. R. Miller, C. Shi, J. T. Tarrasch, J. A. Sundlov, C. Leigh Allen, G. Skiniotis, C. C. Aldrich and A. M. Gulick, *Nature*, 2016, **529**, 235–238.

[25] V. Siewers, R. San-Bento and J. Nielsen, *Biotechnology and Bioengineering*, 2010, **106**, 841–844.

[26] J. Masschelein, C. Clauwers, U. R. Awodi, K. Stalmans, W. Vermaelen, E. Lescrinier, A. Aertsen, C. Michiels, G. L. Challis and R. Lavigne, *Chemical Science*, 2015, **6**, 923–929.

[27] J. K. Stoops, M. J. Arslanian, Y. H. Oh, K. C. Aune, T. C. Vanaman and S. J. Wakil, *Proceedings of the National Academy of Sciences of the United States of America*, 1975, **72**, 1940–1944.

[28] L. Vieweg, S. Reichau, R. Schobert, P. F. Leadlay and R. D. Süssmuth, *Natural Product Reports*, 2014, **31**, 1554–84.

[29] A. Das and C. Khosia, *NIH public access*, 2010, **42**, 631–639.

[30] D. Yu, F. Xu, J. Zeng and J. Zhan, *IUBMB life*, 2012, **64**, 285–95.

[31] L. Katz, *Chemical Reviews*, 1997, **97**, 2557–2576.

[32] G. M. Cragg and D. J. Newman, *Biochimica et Biophysica Acta*, 2013, **1830**, 3670–3695.

[33] B. D. Ames, C. Nguyen, J. Bruegger, P. Smith, W. Xu, S. Ma, E. Wong, S. Wong, X. Xie, J. W. H. Li, J. C. Vederas, Y. Tang and S. Tsai, *Proceedings of the National Academy of Sciences of the United States of America*, 2012, **109**, 11144–11149.

[34] M. Sajish and P. Schimmel, *Nature*, 2015, **519**, 370–373.

[35] H. Brockmann and W. Henkel, *Chemische Berichte*, 1951, **VI**, 284–288.

[36] Y. Xue and D. H. Sherman, *Nature*, 2000, **403**, 571–575.

[37] L. Song, M. Jenner, J. Masschelein, C. Jones, M. J. Bull, S. R. Harris, R. C. Hartkoorn, A. Vocat, I. Romero-Canelon, P. Coupland, G. Webster, M. Dunn, R. Weiser, C. Paisey, S. T. Cole, J. Parkhill, E. Mahenthiralingam and G. L. Challis, *Journal of the American Chemical Society*, 2017, **139**, 7974–7981.

[38] J. Staunton, *Angewandte Chemie International Edition in English*, 1991, **30**, 1302–1306.

[39] E. Mahenthiralingam, T. A. Urban and J. B. Goldberg, *Nature reviews. Microbiology*, 2005, **3**, 144–56.

[40] H. Rose, A. Baldwin, C. G. Dowson and E. Mahenthiralingam, *The Journal of antimicrobial chemotherapy*, 2009, **63**, 502–10.

[41] J. R. Whicher, S. S. Smaga, D. A. Hansen, W. C. Brown, W. H. Gerwick, D. H. Sherman and J. L. Smith, 2014, **20**, year.

[42] Y. A. Chan, A. M. Podevels, B. M. Kevany and M. G. Thomas, *Natural Product Reports*, 2009, **26**, 90–114.

[43] X. Xie, A. Garg, A. T. Keatinge-Clay, C. Khosla and D. E. Cane, *Biochemistry*, 2016, **55**, 1179–1186.

[44] E. Rodriguez, H. G. Menzella and H. Gramajo, *Methods in enzymology*, 2009, **459**, 339–65.

[45] Y. Q. Cheng, J. M. Coughlin, S. K. Lim and B. Shen, *Type I polyketide synthases that require discrete acyltransferases*, Elsevier Inc., 1st edn., 2009, vol. 459, pp. 165–86.

[46] C. Hertweck, A. Luzhetskyy, Y. Rebets and A. Bechthold, *Natural Product Reports*, 2007, **24**, 162–190.

[47] P. Javidpour, C. K. Abhirup Das and S. C. Tsai, *Biochemistry*, 2011, **50**, 7426–7439.

[48] Y. Katsuyama and Y. Ohnishi, *Methods in enzymology*, 2012, **515**, 359–77.

[49] Z. Zhang, H.-X. Pan and G.-L. Tang, *F1000 Research*, 2017, **6**, 1–12.

[50] a. D. Butt and J. B. Ohlrogge, *Plant physiology*, 1991, **96**, 937–942.

[51] M. L. Ernst-Fonberg, M. M. G. Tucker and I. B. Fonberg, *FEBS Letters*, 1987, **215**, 261–265.

[52] D. Gully, D. Moinier, L. Loiseau, E. Bouveret and J. Aiguier, *FEBS Letters*, 2003, **548**, 90–96.

[53] T. A. Holak, S. K. Kearsley, Y. Kim and J. H. Prestegard, *Biochemistry*, 1988, **27**, 6135–6142.

[54] K. A. McAllister, R. B. Peery and G. Zhao, *Journal of bacteriology*, 2006, **188**, 4737–48.

[55] Y. Kim and J. H. Prestegard, *Biochemistry*, 1989, **28**, 8792–8797.

[56] J. Crosby and M. P. Crump, *Nat. Prod. Rep.*, 2012, 1111–1137.

[57] S. A. Newmister and D. H. Sherman, *Chembiochem : A European journal of chemical biology*, 2014, **15**, 1079–81.

[58] R. Bunet, R. Riclea, L. Laureti, L. Hôtel, C. Paris, J.-M. Girardet, D. Spiteller, J. S. Dickschat, P. Leblond and B. Aigle, *PloS one*, 2014, **9**, e87607.

[59] D. I. Chan, T. Stockner, D. P. Tieleman and H. J. Vogel, *The Journal of Biological Chemistry*, 2008, **283**, 33620–9.

[60] A. K. Joshi, A. Witkowski, H. A. Berman, L. Zhang and S. Smith, *Biochemistry*, 2005, **44**, 4100–4107.

[61] L. E. Quadri, P. H. Weinreb, M. Lei, M. M. Nakano, P. Zuber and C. T. Walsh, *Biochemistry*, 1998, **37**, 1585–95.

[62] K. C. Strickland, L. A. Hoeferlin, N. V. Oleinik, N. I. Krupenko and S. A. Krupenko, *The Journal of biological chemistry*, 2010, **285**, 1627–33.

[63] M. A. Skiba, A. P. Sikkema, N. A. Moss, A. N. Lowell, M. Su, R. M. Sturgis, L. Gerwick, W. H. Gerwick, D. H. Sherman and J. L. Smith, *ACS Chemical Biology*, 2018, **13**, 1640–1650.

[64] T. Brautaset, S. E. F. Borgos, H. Sletta, T. E. Ellingsen and S. B. Zotchev, *Journal of Biological Chemistry*, 2003, **278**, 14913–14919.

[65] F. Wang, Y. Wang, J. Ji, Z. Zhou, J. Yu, H. Zhu, Z. Su, L. Zhang and J. Zheng, *ACS Chemical Biology*, 2015, **10**, 1017–1025.

[66] Y. Huang, G. L. Tang, G. Pan, C. Y. Chang and B. Shen, *Organic Letters*, 2016, **18**, 4288–4291.

[67] B. J. Dunn, D. E. Cane and C. Khosla, 2013, 1839–1841.

[68] D. C. Gay, G. Gay, A. J. Axelrod, M. Jenner, C. Kohlhaas, A. Kampa, N. J. Oldham, J. Piel and A. T. Keatinge-Clay, *Structure (London, England : 1993)*, 2014, **22**, 444–51.

[69] T. Robbins, J. Kapilivsky, D. E. Cane and C. Khosla, *Biochemistry*, 2016, acs.biochem.6b00639.

[70] T. Robbins, Y.-C. Liu, D. E. Cane and C. Khosla, *Current Opinion in Structural Biology*, 2016, **41**, 10–18.

[71] A. T. Keatinge-Clay and R. M. Stroud, *Structure*, 2006, **14**, 737–748.

[72] A. T. Keatinge-Clay, *Chemistry & Biology*, 2007, **14**, 898–908.

[73] J. Zheng, S. K. Piasecki and A. T. Keatinge-Clay, *ACS Chemical Biology*, 2016, **8**, 1922–2013.

[74] A. Keatinge-Clay, *Journal of Molecular Biology*, 2008, **384**, 941–953.

[75] D. L. Akey, J. R. Razelun, J. Tehranisa, D. H. Sherman, W. H. Gerwick and J. L. Smith, *Structure*, 2010, **18**, 94–105.

[76] J. Zheng, D. C. Gay, B. Demeler, M. A. White and A. T. Keatinge-Clay, *Nature Chemical Biology*, 2013, **8**, 615–621.

[77] M. A. Skiba, A. P. Sikkema, W. D. Fiers, W. H. Gerwick, D. H. Sherman, C. C. Aldrich and J. L. Smith, *ACS Chemical Biology*, 2016, **11**, 3319–3327.

[78] M. Jenner, J. P. Afonso, H. R. Bailey, S. Frank, A. Kampa, J. Piel and N. J. Oldham, *Angewandte Chemie - International Edition*, 2015, **54**, 1817–1821.

[79] M. Jenner, J. Afonso, H. R. Bailey, S. Frank, A. Kampa, J. Piel and N. J. Oldham, *Angewandte Chemie*, 2015, 1–86.

[80] A. Rittner and M. Grininger, *ChemBioChem*, 2014, **15**, 2489–2493.

[81] H. Chen and L. C. Du, *Applied & Environmental Microbiology*, 2016, **100**, 541–557.

[82] Y. Abe, T. Suzuki, C. Ono, K. Iwamoto, M. Hosobuchi and H. Yoshikawa, *Molecular Genetics & Genomics*, 2002, **267**, 636–646.

[83] K. Kasahara, T. Miyamoto, T. Fujimoto, H. Oguri, T. Tokiwano, H. Oikawa, Y. Ebizuka and I. Fujii, *Chembiochem : A European journal of chemical biology*, 2010, **11**, 1245–52.

[84] P. Caffrey, D. J. Bevitt, J. Staunton and P. F. Leadlay, *FEBS Letters*, 1992, **304**, 225–228.

[85] B. Shen, *Current Opinion in Chemical Biology*, 2003, **7**, 285–295.

[86] J. F. Aparicio, P. Caffrey, A. F. A. Marsden, J. Staunton and P. F. Leadlay, *Journal of Biological Chemistry*, 1994, **269**, 8524–8528.

[87] G. A. Roberts, J. Staunton and P. F. Leadlay, *European Journal of Biochemistry*, 1993, **214**, 305–311.

[88] J. Cortes, K. Wiesmann, G. Roberts, M. Brown, J. Staunton and P. Leadlay, *Science*, 1995, **268**, 1487–1489.

[89] L. K. Charkoudian, C. W. Liu, S. Capone, S. Kapur, D. E. Cane, A. Togni, D. Seebach and C. Khosla, *Protein Science*, 2011, **20**, 1244–1255.

[90] A. L. Edwards, T. Matsui, T. M. Weiss and C. Khosla, *Journal of Molecular Biology*, 2014, **426**, 2229–2245.

[91] J. Piel, *Natural Product Reports*, 2010, **27**, 996–1047.

[92] F. T. Wong, X. Jin, I. I. Mathews, D. E. Cane and C. Khosla, *Biochemistry*, 2011, **50**, 6539–6548.

[93] E. Mahenthiralingam, L. Song, A. Sass, J. White, C. Wilmot, A. Marchbank, O. Boaisha, J. Paine, D. Knight and G. L. Challis, *Chemistry & Biology*, 2011, **18**, 665–77.

[94] N. Wu, D. E. Cane and C. Khosla, *Biochemistry*, 2002, **41**, 5056–5066.

[95] G. Yadav, R. S. Gokhale and D. Mohanty, *PLoS Computational Biology*, 2009, **5**, 1–14.

[96] B. I. Khayatt, L. Overmars, R. J. Siezen and C. Francke, *PloS one*, 2013, **8**, e62136.

[97] D. H. Kwan, Y. Sun, F. Schulz, H. Hong, B. Popovic, J. C. Sim-Stark, S. F. Haydock and P. F. Leadlay, *Chemistry & Biology*, 2008, **15**, 1231–1240.

[98] S. Khater, M. Gupta, P. Agrawal, N. Sain, J. Prava, P. Gupta, M. Grover, N. Kumar and D. Mohanty, *Nucleic Acids Research*, 2017, **45**, W72–W79.

[99] Kumar P, K. C and T. Y, *Methods Enzymol*, 2004, **388**, 269–93.

[100] Q. Wu, J. Liang, S. Lin, X. Zhou, L. Bai, Z. Deng and Z. Wang, *Antimicrobial Agents & Chemotherapy*, 2011, **55**, 974–982.

[101] J. Zhang, S. G. Van Lanen, J. Ju, W. Liu, P. C. Dorrestein, W. Li, N. L. Kelleher and B. Shen, *Proceedings of the National Academy of Sciences of the United States of America*, 2008, **105**, 1460–1465.

[102] B. J. Beck, Y. J. Yoon, K. A. Reynolds and D. H. Sherman, *The Hidden Steps of Domain Skipping*, 2002.

[103] J. D. Kittendorf and D. H. Sherman, *Bioorg Med Chem.*, 2010, **17**, 2137–2146.

[104] B. Beck, Y. Yoon, K. Reynolds and D. Sherman, *Chem. Biol*, 2002, **9**, 575–583.

[105] J. Brink, S. J. Ludtke, C. Y. Yang, Z. W. Gu, S. J. Wakil and W. Chiu, *Proceedings of the National Academy of Sciences of the United States of America*, 2002, **99**, 138–43.

[106] M. Leibundgut, T. Maier, S. Jenni and N. Ban, *Current Opinion in Structural Biology*, 2008, 714–725.

[107] S. Smith and S.-c. Tsai, *Nat. Prod. Rep.*, 2008, **24**, 1041–1072.

[108] Y. Tang, C. Y. Kim, I. I. Mathews, D. E. Cane and C. Khosla, *Proceedings of the National Academy of Sciences of the United States of America*, 2006, **103**, 11124–11129.

[109] S. C. Tsai, L. J. W. Miercke, J. Krucinski, R. Gokhale, J. C.-H. Chen, P. G. Foster, D. E. Cane, C. Khosla and R. M. Stroud, *Proceedings of the National Academy of Sciences of the United States of America*, 2001, **98**, 14808–14813.

[110] S. K. Piasecki, J. Zheng, A. J. Axelrod, M. E. Detelich and A. T. Keatinge-Clay, *Proteins: Structure, Function & Bioinformatics*, 2014, **82**, 2067–2077.

[111] E. L. Noey, N. Tibrewal, G. Jiménez-Osés, S. Osuna, J. Park, C. M. Bond, D. Cascio, J. Liang, X. Zhang, G. W. Huisman, Y. Tang and K. N. Houk, *Proceedings of the National Academy of Sciences of the United States of America*, 2015, 201507910.

[112] W. D. Fiers, G. J. Dodge, D. H. Sherman, J. L. Smith and C. C. Aldrich, *Journal of the American Chemical Society*, 2016, **138**, 16024–16036.

[113] D. A. Herbst, R. P. Jakob, F. Zähringer and T. Maier, *Nature*, 2016, **531**, 533–537.

[114] T. P. Korman, J. M. Crawford, J. W. Labonte, A. G. Newman, J. Wong, C. A. Townsend and S. C. Tsai, *Proceedings of the National Academy of Sciences of the United States of America*, 2010, **107**, 6246–6251.

[115] P. Argyropoulos, F. Bergeret, C. Pardin, J. M. Reimer, A. Pinto, C. N. Boddy and T. M. Schmeing, *Biochimica et Biophysica Acta - General Subjects*, 2016, **1860**, 486–497.

[116] Y. Tang, A. Y. Chen, C. Y. Kim, D. E. Cane and C. Khosla, *Chemistry & Biology*, 2007, **14**, 931–943.

[117] J. Whicher, *Ph.D. thesis*, University of Michigan, 2013.

[118] L. Schr\"odinger, *The PyMOL Molecular Graphics System, Version 1.8*, 2015.

[119] K. Arakawa, *Nature Chemical Biology*, 2012, **8**, 604–605.

[120] C. H. Eng, T. W. H. Backman, C. B. Bailey, C. Magnan, L. Katz, P. Baldi and J. D. Keasling, *Nucleic Acids Research*, 2018, **46**, 509–515.

[121] A. T. Keatinge-Clay, *Natural Product Reports*, 2012, **29**, 1050.

[122] J. Zheng, C. D. Fage, B. Demeler, D. W. Hoffman and A. T. Keatinge-Clay, *ACS Chemical Biology*, 2013, **8**, 1263–1270.

[123] S. Dutta, J. R. Whicher, D. A. Hansen, W. A. Hale, J. A. Chemler, G. R. Congdon, A. R. H. Narayan, K. Håkansson, D. H. Sherman, J. L. Smith and G. Skiniotis, *Nature*, 2014, **510**, 512–7.

[124] R. Broadhurst, D. Nietlispach, P. Wheatcroft, P. F. Leadlay and K. J. Weissman, *Chemistry & Biology*, 2003, **10**, 723–731.

[125] T. J. Buchholz, T. W. Geders, F. E. Bartley III, K. A. Reynolds, L. Janet and D. H. Sherman, *Chemical Biology*, 2009, **4**, 41–52.

[126] J. R. Whicher, S. S. Smaga, D. A. Hansen, W. C. Brown, W. H. Gerwick, D. H. Sherman and J. L. Smith, *Chemistry & Biology*, 2013, **20**, 1340–1351.

[127] L. Gu, E. B. Eisman, S. Dutta, T. M. Franzmann, S. Walter, W. H. Gerwick, G. Skiniotis and D. H. Sherman, *Angewandte Chemie - International Edition*, 2011, **50**, 2795–2798.

[128] J. Dorival, T. Annaval, F. Risser, S. Collin, P. Roblin, C. Jacob, A. Gruez, B. Chagot and K. J. Weissman, *Journal of the American Chemical Society*, 2016, **138**, 4155–4167.

[129] M. Thattai, Y. Burak and B. I. Shraiman, *PLoS Computational Biology*, 2007, **3**, 1827–1835.

[130] M. Klaus, M. P. Ostrowski, J. Austerjost, T. Robbins, B. Lowry, D. E. Cane and C. Khosla, *Journal of Biological Chemistry*, 2016, jbc.M116.730531.

[131] D. Morikis, M. Roy, M. G. Newlon, J. D. Scott and P. A. Jennings, *European Journal of Biochemistry*, 2002, **269**, 2040–2051.

[132] P. Kumar, Q. Li, D. E. Cane and C. Khosla, *Journal of the American Chemical Society*, 2003, **125**, 4097–4102.

[133] A. Parmeggiani, I. M. Krab, T. Watanabe, R. C. Nielsen, C. Dahlberg, J. Nyborg and P. Nissen, *The Journal of Biological Chemistry*, 2006, **281**, 2893–900.

[134] C. D. Richter, D. Nietlispach, R. W. Broadhurst and K. J. Weissman, *Nature Chemical Biology*, 2008, **4**, 75–81.

[135] J. Staunton and K. J. Weissman, *Natural Product Reports*, 2001, **18**, 380–416.

[136] H.-Q. Pan, S.-Y. Zhang, N. Wang, Z.-L. Li, H.-M. Hua, J.-C. Hu and S.-J. Wang, *Marine drugs*, 2013, **11**, 3891–901.

[137] M. H. Lacoske and E. A. Theodorakis, *Journal of Natural Products*, 2015, **78**, 562–575.

[138] P. W. Clutterbuck, H. Raistrick and F. Reuter, *Biochemical Journal*, 1935, **29**, 1300–1309.

[139] H. Roh, G. C. Uguru, H. J. Ko, S. Kim, B. Y. Kim, M. Goodfellow, A. T. Bull, K. H. Kim, M. J. Bibb, I. G. Choi and J. E. M. Stach, *Journal of Bacteriology*, 2011, **193**, 3391–3392.

[140] J. S. Freundlich, M. Lalgondar, J. R. Wei, S. Swanson, E. J. Sorensen, E. J. Rubin and J. C. Sacchettini, *Tuberculosis*, 2010, **90**, 298–300.

[141] A. K. Mallams, M. S. Puar, R. R. Rossman, A. T. McPhail and R. D. Macfarlane, *Journal of the American Chemical Society*, 1981, **103**, 3940–3943.

[142] T. Nakashima, M. Miura and M. Hara, *Cancer Research*, 2000, **60**, 1229–1235.

[143] T. Tsunakawa, O. Tenmyo, K. Tomita, N. Naruse, K. Chikako, T. Miyaki, M. Konishi and T. Oki, *The Journal of Antibiotics*, 1992, **45**, 180–188.

[144] P. M. Small and H. F. Chambers, *Antimicrobial Agents and Chemotherapy*, 1990, **34**, 1227–1231.

[145] D. J. McGeoch, F. J. Rixon and A. J. Davison, *Virus Research*, 2006, **117**, 90–104.

[146] A. Tanabe-Tochikura, H. Nakashima, T. Murakami, O. Tenmyo, T. Oki and N. Yamamoto, *Antiviral Chemistry & Chemotherapy*, 1992, **3**, 345–349.

[147] H. Y. He, H. X. Pan, L. F. Wu, B. B. Zhang, H. B. Chai, W. Liu and G. L. Tang, *Chemistry & Biology*, 2012, **19**, 1313–23.

[148] R. Gibson, *Ph.D. thesis*, 2014.

[149] M. Zerikly and G. L. Challis, *ChemBioChem*, 2009, **10**, 625–633.

[150] W. Roush and D. Barda, *Tetrahedron letters*, 1997, **38**, 8785–8788.

[151] S. Poust, I. Yoon, P. D. Adams, L. Katz, C. J. Petzold and J. D. Keasling, *PLoS ONE*, 2014, **9**, 2–5.

[152] E. M. Gottardi, J. M. Krawczyk, H. von Suchodoletz, S. Schadt, A. Mühlenweg, G. C. Uguru, S. Pelzer, H. P. Fiedler, M. J. Bibb, J. E. M. Stach and R. D. Süssmuth, *ChemBioChem*, 2011, **12**, 1401–1410.

[153] J. S. Freundlich, M. Lalgondar, J. R. Wei, S. Swanson, E. J. Sorensen, E. J. Rubin and J. C. Sacchettini, *Tuberculosis*, 2010, **90**, 298–300.

[154] Y. Sun, H. Hong, F. Gillies, J. B. Spencer and P. F. Leadlay, *ChemBioChem*, 2008, **9**, 150–156.

[155] L. F. Wu, H. Y. He, H. X. Pan, L. Han, R. Wang and G. L. Tang, *Organic Letters*, 2014, **16**, 1578–1581.

[156] W. R. Roush, D. A. Barda, C. Limberakis and R. K. Kunz, 2002, **58**, 6433–6454.

[157] M. K. Zerikly and G. L. Challis, *Ph.D. thesis*, the University of Warwick, 2010.

[158] D. Meluzzi, W. H. Zheng, M. Hensler, V. Nizet and P. C. Dorrestein, *Bioorg Med Chem.*, 2008, **18**, 3107–3111.

[159] H. Liu, H. Zhang, J. D. King, N. R. Wolf, M. Prado, M. L. Gross and R. E. Blankenship, *Biochimica et Biophysica Acta - Bioenergetics*, 2014, **1837**, 1955–1963.

[160] J. Bruegger, B. Haushalter, A. Vagstad, G. Shakya, N. Mih, C. A. Townsend, M. D. Burkart and S. C. Tsai, *Chemistry & Biology*, 2013, **20**, 1135–1146.

[161] N. M. Kosa, K. M. Pham and M. D. Burkart, *Chemical Science*, 2014, **5**, 1179.

[162] L. Schmohl and D. Schwarzer, *Journal of peptide science : an official publication of the European Peptide Society*, 2014, 145–151.

[163] R. M. Kaake, X. Wang, A. Burke, C. Yu, W. Kandur, E. J. Novtisky, T. Second, J. Duan, A. Kao and S. Guan, 2014.

[164] S. E. Klohr, D. R. Schroeder, K. L. Colson, M. S. Lee, J. a. Matson, L. S. Brinen and J. Clardy, **49**, year.

[165] A. Doerr, *Nature Methods*, 2013, **11**, 30–30.

[166] A. Miyanaga, *Proceedings of the National Academy of Sciences of the United States of America*, 2016, **113**, 1802–7.

[167] S. Mark and C. I. Laboratories.

[168] M. M. Bradford, *Analytical Biochemistry*, 1976, **72**, 248–254.

[169] M. Yamashita and J. B. Fenn, *J Phys Chem*, 1984, **434**, 4451–4459.

[170] K. F. Geoghegan, H. B. Dixon, P. J. Rosner, L. R. Hoth, A. J. Lanzetti, K. A. Borzilleri, E. S. Marr, L. H. Pezzullo, L. B. Martin, P. K. LeMotte, A. S. McColl, A. V. Kamath and J. G. Stroh, *Analytical biochemistry*, 1999, **267**, 169–84.

[171] M. V. Rubio, M. P. Zubieta, J. P. L. Franco Cairo, F. Calzado, A. F. Paes Leme, F. M. Squina, R. A. Prade and A. R. de Lima Damásio, *Biotechnology for Biofuels*, 2016, **9**, 168.

[172] P. C. Dorrestein, S. B. Bumpus, C. T. Calderone, S. Garneau-Tsodikova, Z. D. Aron, P. D. Straight, R. Kolter, C. T. Walsh and N. L. Kelleher, *ACS Chemical Biology*, 2006, **45**, 12756–12766.

[173] H. Petković, A. Sandmann, I. R. Challis, H.-J. Hecht, B. Silakowski, L. Low, N. Beeston, E. Kuščer, J. Garcia-Bernardo, P. F. Leadlay, S. G. Kendrew, B. Wilkinson and R. Müller, *Organic and Biomolecular Chemistry*, 2008, **6**, 500–506.

[174] K. Watanabe, C. C. Wang, C. N. Boddy, D. E. Cane and C. Khosla, *Journal of Biological Chemistry*, 2003, **278**, 42020–42026.

[175] H. Zhou, J. Zhan, K. Watanabe, X. Xie and Y. Tang, *Proceedings of the National Academy of Sciences of the United States of America*, 2008, **105**, 6249–54.

[176] S. Kapur, A. Worthington, Y. Tang, D. E. Cane, M. D. Burkart and C. Khosla, *Bioorganic & Medicinal Chemistry*, 2008, **18**, 3034–3038.

[177] C. Nguyen, R. W. Haushalter, D. J. Lee, P. R. L. Markwick, J. Bruegger, G. Caldara-Festin, K. Finzel, D. R. Jackson, F. Ishikawa, B. O'Dowd, J. A. McCammon, S. J. Opella, S. C. Tsai and M. D. Burkart, *Nature*, 2014, **505**, 427–31.

[178] A. Miyanaga, S. Iwasawa, Y. Shinohara, F. Kudo and T. Eguchi, *Proceedings of the National Academy of Sciences of the United States of America*, 2016, **113**, 1802–1807.

[179] S. Vance, O. Tkachenko, B. Thomas, M. Bassuni, H. Hong, D. Nietlispach and W. Broadhurst, *Biochemical Journal*, 2016, **473**, 1097–1110.

[180] S. Paul, H. Ishida, L. T. Nguyen, Z. Liu and H. J. Vogel, *Protein Science*, 2017, **26**, 946–959.

[181] Y. G. Park, M. C. Jung, H. Song, K. W. Jeong, E. Bang, G. S. Hwang and Y. Kim, *Journal of Biological Chemistry*, 2016, **291**, 1692–1702.

[182] V. Y. Alekseyev, C. W. Liu, D. E. Cane, J. D. Puglisi and C. Khosla, *Protein Science*, 2007, **16**, 2093–2107.

[183] L. Manzi, A. S. Barrow, D. Scott, R. Layfield, T. G. Wright, J. E. Moses and N. J. Oldham, *Nature Communications*, 2016, **7**, 1–9.

[184] T. D. Colby, B. J. Bahnson, J. K. Chin, J. P. Klinman and B. M. Goldstein, *Biochemistry*, 1998, **37**, 9295–9304.

[185] M. Allaire, Y. Li, R. E. MacKenzie and M. Cygler, *Structure*, 1998, **6**, 173–182.

[186] X. Y. Jia, Z. H. Tian and L. Shao, *Chemistry & Biology*, 2006, **13**, 575–585.

[187] M. J. Bertin, A. Vulpanovici, E. A. Monroe, A. Korobeynikov, D. H. Sherman, L. Gerwick and W. G. H., *ChemBioChem*, 2016, **17**, 164–173.

[188] J. R. Whicher, S. Dutta, D. A. Hansen, W. A. Hale, J. A. Chemler, A. M. Dosey, A. R. H. Narayan, K. Håkansson, D. H. Sherman, J. L. Smith and G. Skiniotis, *Nature*, 2014, **510**, 560–4.

[189] S. Sabban, *Ph.D. thesis*, 2011.

[190] B. Lowry, X. Li, T. Robbins, D. E. Cane and C. Khosla, *ACS Central Science*, 2016, **2**, 14–20.

[191] F. W. Studier and B. A. Moffatt, *Journal of Molecular Biology*, 1986, **189**, 113–130.

[192] F. Bloch, W. W. Hansen and M. Packard, *Physical Review*, 1946, **70**, 474–485.

[193] E. M. Purcell, H. C. Torrey and R. V. Pound, *Resonance Absorption by Nucelar Magnetic Moments in a Solid*, 1946.

[194] P. Atkins, J. de Paula and R. Friedman, *Physical Chemistry: Quanta, Matter, and Change*, Oxford University Press, 2010.

[195] L. A. O'Dell, *Solid State Nuclear Magnetic Resonance*, 2013, **55-56**, 28–41.

[196] W. Lee, M. Tonelli and J. L. Markley, *Bioinformatics*, 2015, **31**, 1325–1327.

[197] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, A. Cardona, K. Liceiri, P. Tomancak and C. A., *Nature Methods*, 2012, **9**, 676–682.

[198] K. Eliceiri, C. A. Schneider, W. S. Rasband and K. W. Eliceiri, *Nature Methods*, 2012, **9**, 671–675.

Chapter 8

Appendix

Constructs

All constructs contain the following N-terminal **His-Tag**: MHHHHHHGKPIPNPLLGLDSTENLYFQGIDPFT

QMN M3_ACP (Dr. Orestis Lazos)

LLDLVTTTVAEVLGHAGTDGIDADRGLLAMGFDSLTAVEFRNRLAAETGLRLPSTLVFNHPTPGAIAEFL
RAQLAPEPTDELDRLAADLSTMDPERQAAVLGRMAELLHKWNTNGTGNGASADLDSLSDEELFDALD
EELGQR

QMN M3_ACPS

LLDLVTTTVAEVLGHAGTDGIDADRGLLAMGFDSLTAVEFRNRLAAETGLRLPSTLVFNHPTPGAIAEFL
RAQLAPEPTD

QMN M4_ACP (Dr. Orestis Lazos)

LLIDVVRGHVATVLGHTDPDTIPTEKAFKELGFDSLTAVELRNRLTAATGLRLPATLVFDHPTLAGLAGYM
RAELAPRTEEPDPASAVLERLAELEEALAGLSDPRIADRLRELLAVAGGTDPGLGSASADELFDFIDGKLG
RKVETA

QMN M4_ACPS

LLIDVVRGHVATVLGHTDPDTIPTEKAFKELGFDSLTAVELRNRLTAATGLRLPATLVFDHPTLAGLAGYM
RAELAP

QMN M5_ACP (residues 1199-1292 of M5)

ATGAEQEQLLTELVRFHASDVLGHASPEAIGVEDSLLDHGFSSFAALELSNRLQAASGVQIPPVAIYDQPT
LGAVITHLRDALAAEPS

QMN M5_KS (residues 1-444 of M5)

A mutation was carried out on the KS active-site (C207A).

MSTGPTEQKLLDYLKWVTTDLDRTRDQLLALEEKASEPIAVVAMSCRYPHGVTSPEELWELVASGRDAIA
DFPADRGWPADVYDPDPDRRDRPHAGGFLYDAADFDPGFFGISPREAPAIDPQQRLLLEIAWEAFERAGI
DAESVRGSDTGVFTGVMYGDYGSRLQDRAPQGYESYLGNGSAGSIASGRVAYTFGLEGPAVTIDTA**C**SSSL
VALHLAVQALRQGDCSLALAGGVTVMATPMVFTEFDRQRGLSADGRCKSFAAAADGTGFAEGAGLLLL
ERLSDARRHGHPVLGLVRGTAVNQDGASNGLTAPNGPSQERVIKQALENARVTPSEVDAVEAHGTGTTL
GDPIEAQALLATYGKARGEKPLYLGSIKSNIGHTQAAAGVAGVIKMIMAMRHGELPRTLHVDAPNPLVD
WASGGVSLLTEPAKWQQNGHPRRAGVSSFGISGTNAHVILEQAPPVE

QMN M5_KSATS (residues 1-888 of M5)

Mutations were carried out on the KS active-site (C207A) and the AT active-site (S656C).

MSTGPTEQKLLDYLKWVTTDLDRTRDQLLALEEKASEPIAVVAMSCRYPHGVTSPEELWELVASGRDAIA
DFPADRGWPADVYDPDPDRRDRPHAGGFLYDAADFDPGFFGISPREAPAIDPQQRLLLEIAWEAFERAGI
DAESVRGSDTGVFTGVMYGDYGSRLQDRAPQGYESYLGNGSAGSIASGRVAYTFGLEGPAVTIDTA**C**SSSL
VALHLAVQALRQGDCSLALAGGVTVMATPMVFTEFDRQRGLSADGRCKSFAAAADGTGFAEGAGLLLL
ERLSDARRHGHPVLGLVRGTAVNQDGASNGLTAPNGPSQERVIKQALENARVTPSEVDAVEAHGTGTTL
GDPIEAQALLATYGKARGEKPLYLGSIKSNIGHTQAAAGVAGVIKMIMAMRHGELPRTLHVDAPNPLVD
WASGGVSLLTEPAKWQQNGHPRRAGVSSFGISGTNAHVILEQAPPVEPASEPCVAPGPVPWVLSAKSPAA
VQGQAERLLAHLTLNPGLTVADVGHSLATTRTSFSHRAVVFGDRQEELVKGLSGLAHGTPATNLVQGRAT
GGKTAFVFTGQGSQRPGMGRQLYATYPVFTAELDRVCALFDGLLDRPLLPILFAEPRTPEAALLNRTAYTQ
VALFALETALYRLVESYRVRPSFVAGH**S**IGELTAAHVAGVLSLEDACVLVAARGRLMESARDDGAMLAIEA
PESAVTPLLDASVSLAAVNGPNAVVLAGDLERVLAIGAEFKAQGVRTKRLTVSHAFHSAHLDPMLDEYRR
IARGLSYSAPRIPVVSTVTGEVADELTSPEYWVRQVRQPVRFLDGMRTLYSSGVRTFVEIGPDAVLAPAMQA
CVPEVDHGVIPALTSRRSEPRTLVTALAKAHVRGVPVDWSAGFSGAR

QMN_KSATΔ283 (residues 1-915 of M5)

MSTGPTEQKLLDYLKWVTTDLDRTRDQLLALEEKASEPIAVVAMSCRYPHGVTSPEELWELVASGRDAIA
DFPADRGWPADVYDPDPDRRDRPHAGGFLYDAADFDPGFFGISPREAPAIDPQQRLLLEIAWEAFERAGI
DAESVRGSDTGVFTGVMYGDYGSRLQDRAPQGYESYLGNGSAGSIASGRVAYTFGLEGPAVTIDTA**C**SSSL
VALHLAVQALRQGDCSLALAGGVTVMATPMVFTEFDRQRGLSADGRCKSFAAAADGTGFAEGAGLLLL

ERLSDARRHGHPVLGLVRGTAVNQDGASNGLTAPNGPSQERVIKQALENARVTPSEVDAVEAHGTGTTL
GDPIEAQALLATYGKARGEKPLYLGSIKSNIGHTQAAAGVAGVIKMIMAMRHGELPRTLHVDAPNPLVD
WASGGVSLLTEPAKWQQNGHPRRAGVSSFGISGTNAHVILEQAPPVEPASEPCVAPGPVPWVLSAKSPAA
VQGQAERLLAHLTLNPGLTVADVGHSLATTRTSFSHRAVVFGDRQEELVKGLSGLAHGTPATNLVQGRAT
GGKTAFVFTGQGSQRPGMGRQLYATYPVFTAELDRVCALFDGLLDRPLLPILFAEPRTPEAALLNRTAYTQ
VALFALETALYRLVESYRVRPSFVAGH**S**IGELTAAHVAGVLSLEDACVLVAARGRLMESARDDGAMLAIEA
PESAVTPLLDASVSLAAVNGPNAVVLAGDLERVLAIGAEFKAQGVRTKRLTVSHAFHSAHLDPMLDEYRR
IARGLSYSAPRIPVVSTVTGEVADELTSPEYWVRQVRQPVRFLDGMRTLYSSGVRTFVEIGPDAVLAPAMQA
CVPEVDHGVIPALTSRRSEPRTLVTALAKAHVRGVPVDWSAGFSGARRVDLPTYAFQRSRYWLDPEAPSAQ
VD

QMN_KSATΔ136 (residues 1-1062 of M5)

MSTGPTEQKLLDYLKWVTTDLDRTRDQLLALEEKASEPIAVVAMSCRYPHGVTSPEELWELVASGRDAIA
DFPADRGWPADVYDPDPDRRDRPHAGGFLYDAADFDPGFFGISPREAPAIDPQQRLLLEIAWEAFERAGI
DAESVRGSDTGVFTGVMYGDYGSRLQDRAPQGYESYLGNGSAGSIASGRVAYTFGLEGPAVTIDTA**C**SSSL
VALHLAVQALRQGDCSLALAGGVTVMATPMVFTEFDRQRGLSADGRCKSFAAAADGTGFAEGAGLLLL
ERLSDARRHGHPVLGLVRGTAVNQDGASNGLTAPNGPSQERVIKQALENARVTPSEVDAVEAHGTGTTL
GDPIEAQALLATYGKARGEKPLYLGSIKSNIGHTQAAAGVAGVIKMIMAMRHGELPRTLHVDAPNPLVD
WASGGVSLLTEPAKWQQNGHPRRAGVSSFGISGTNAHVILEQAPPVEPASEPCVAPGPVPWVLSAKSPAA
VQGQAERLLAHLTLNPGLTVADVGHSLATTRTSFSHRAVVFGDRQEELVKGLSGLAHGTPATNLVQGRAT
GGKTAFVFTGQGSQRPGMGRQLYATYPVFTAELDRVCALFDGLLDRPLLPILFAEPRTPEAALLNRTAYTQ
VALFALETALYRLVESYRVRPSFVAGH**S**IGELTAAHVAGVLSLEDACVLVAARGRLMESARDDGAMLAIEA
PESAVTPLLDASVSLAAVNGPNAVVLAGDLERVLAIGAEFKAQGVRTKRLTVSHAFHSAHLDPMLDEYRR
IARGLSYSAPRIPVVSTVTGEVADELTSPEYWVRQVRQPVRFLDGMRTLYSSGVRTFVEIGPDAVLAPAMQA
CVPEVDHGVIPALTSRRSEPRTLVTALAKAHVRGVPVDWSAGFSGARRVDLPTYAFQRSRYWLDPEAPSAQ
VDLASAADVLGLTEVQRATLAAWQRRHTWAHRKTWRPLRTVPSPPGTWLVPSGPDRLVEVLRERGAQVV
TAGNAPVDGVLTFDEVDHPAPVWLLTSDADPGYACDHLVHLPTSPTDETWERLVDALGSAEPRLAIRGPE
VFAEHLERAEL

QMN_KSATΔ72 (residues 1-1126 of M5)

MSTGPTEQKLLDYLKWVTTDLDRTRDQLLALEEKASEPIAVVAMSCRYPHGVTSPEELWELVASGRDAIA
DFPADRGWPADVYDPDPDRRDRPHAGGFLYDAADFDPGFFGISPREAPAIDPQQRLLLEIAWEAFERAGI
DAESVRGSDTGVFTGVMYGDYGSRLQDRAPQGYESYLGNGSAGSIASGRVAYTFGLEGPAVTIDTA**C**SSSL
VALHLAVQALRQGDCSLALAGGVTVMATPMVFTEFDRQRGLSADGRCKSFAAAADGTGFAEGAGLLLL

ERLSDARRHGHPVLGLVRGTAVNQDGASNGLTAPNGPSQERVIKQALENARVTPSEVDAVEAHGTGTTL
GDPIEAQALLATYGKARGEKPLYLGSIKSNIGHTQAAAGVAGVIKMIMAMRHGELPRTLHVDAPNPLVD
WASGGVSLLTEPAKWQQNGHPRRAGVSSFGISGTNAHVILEQAPPVEPASEPCVAPGPVPWVLSAKSPAA
VQGQAERLLAHLTLNPGLTVADVGHSLATTRTSFSHRAVVFGDRQEELVKGLSGLAHGTPATNLVQGRAT
GGKTAFVFTGQGSQRPGMGRQLYATYPVFTAELDRVCALFDGLLDRPLLPILFAEPRTPEAALLNRTAYTQ
VALFALETALYRLVESYRVRPSFVAGHSIGELTAAHVAGVLSLEDACVLVAARGRLMESARDDGAMLAIEA
PESAVTPLLDASVSLAAVNGPNAVVLAGDLERVLAIGAEFKAQGVRTKRLTVSHAFHSAHLDPMLDEYRR
IARGLSYSAPRIPVVSTVTGEVADELTSPEYWVRQVRQPVRFLDGMRTLYSSGVRTFVEIGPDAVLAPAMQA
CVPEVDHGVIPALTSRRSEPRTLVTALAKAHVRGVPVDWSAGFSGARRVDLPTYAFQRSRYWLDPEAPSAQ
VDLASAADVLGLTEVQRATLAAWQRRHTWAHRKTWRPLRTVPSPPGTWLVPSGPDRLVEVLRERGAQVV
TAGNAPVDGVLTFDEVDHPAPVWLLTSDADPGYACDHLVHLPTSPTDETWERLVDALGSAEPRLAIRGPE
VFAEHLERAELDGDWWPDGPVLITGDDELSSWAARWFEEQGVAVVAEAGPSLAAVVQVVPPEWTPAEVR
EALAPL

QMN_KSATΔ60 (residues 1-1138 of M5)

MSTGPTEQKLLDYLKWVTTDLDRTRDQLLALEEKASEPIAVVAMSCRYPHGVTSPEELWELVASGRDAIA
DFPADRGWPADVYDPDPDRRDRPHAGGFLYDAADFDPGFFGISPREAPAIDPQQRLLLEIAWEAFERAGI
DAESVRGSDTGVFTGVMYGDYGSRLQDRAPQGYESYLGNGSAGSIASGRVAYTFGLEGPAVTIDTACSSSL
VALHLAVQALRQGDCSLALAGGVTVMATPMVFTEFDRQRGLSADGRCKSFAAAADGTGFAEGAGLLLL
ERLSDARRHGHPVLGLVRGTAVNQDGASNGLTAPNGPSQERVIKQALENARVTPSEVDAVEAHGTGTTL
GDPIEAQALLATYGKARGEKPLYLGSIKSNIGHTQAAAGVAGVIKMIMAMRHGELPRTLHVDAPNPLVD
WASGGVSLLTEPAKWQQNGHPRRAGVSSFGISGTNAHVILEQAPPVEPASEPCVAPGPVPWVLSAKSPAA
VQGQAERLLAHLTLNPGLTVADVGHSLATTRTSFSHRAVVFGDRQEELVKGLSGLAHGTPATNLVQGRAT
GGKTAFVFTGQGSQRPGMGRQLYATYPVFTAELDRVCALFDGLLDRPLLPILFAEPRTPEAALLNRTAYTQ
VALFALETALYRLVESYRVRPSFVAGHSIGELTAAHVAGVLSLEDACVLVAARGRLMESARDDGAMLAIEA
PESAVTPLLDASVSLAAVNGPNAVVLAGDLERVLAIGAEFKAQGVRTKRLTVSHAFHSAHLDPMLDEYRR
IARGLSYSAPRIPVVSTVTGEVADELTSPEYWVRQVRQPVRFLDGMRTLYSSGVRTFVEIGPDAVLAPAMQA
CVPEVDHGVIPALTSRRSEPRTLVTALAKAHVRGVPVDWSAGFSGARRVDLPTYAFQRSRYWLDPEAPSAQ
VDLASAADVLGLTEVQRATLAAWQRRHTWAHRKTWRPLRTVPSPPGTWLVPSGPDRLVEVLRERGAQVV
TAGNAPVDGVLTFDEVDHPAPVWLLTSDADPGYACDHLVHLPTSPTDETWERLVDALGSAEPRLAIRGPE
VFAEHLERAELDGDWWPDGPVLITGDDELSSWAARWFEEQGVAVVAEAGPSLAAVVQVVPPEWTPAEVR
EALAPLDELTAMVFVFTS

QMN_KSATΔ38 (residues 1-1160 of M5)

MSTGPTEQKLLDYLKWVTTDLDRTRDQLLALEEKASEPIAVVAMSCRYPHGVTSPEELWELVASGRDAIA
DFPADRGWPADVYDPDPDRRDRPHAGGFLYDAADFDPGFFGISPREAPAIDPQQRLLLEIAWEAFERAGI
DAESVRGSDTGVFTGVMYGDYGSRLQDRAPQGYESYLGNGSAGSIASGRVAYTFGLEGPAVTIDTACSSSL
VALHLAVQALRQGDCSLALAGGVTVMATPMVFTEFDRQRGLSADGRCKSFAAAADGTGFAEGAGLLLL
ERLSDARRHGHPVLGLVRGTAVNQDGASNGLTAPNGPSQERVIKQALENARVTPSEVDAVEAHGTGTTL
GDPIEAQALLATYGKARGEKPLYLGSIKSNIGHTQAAAGVAGVIKMIMAMRHGELPRTLHVDAPNPLVD
WASGGVSLLTEPAKWQQNGHPRRAGVSSFGISGTNAHVILEQAPPVEPASEPCVAPGPVPWVLSAKSPAA
VQGQAERLLAHLTLNPGLTVADVGHSLATTRTSFSHRAVVFGDRQEELVKGLSGLAHGTPATNLVQGRAT
GGKTAFVFTGQGSQRPGMGRQLYATYPVFTAELDRVCALFDGLLDRPLLPILFAEPRTPEAALLNRTAYTQ
VALFALETALYRLVESYRVRPSFVAGHSIGELTAAHVAGVLSLEDACVLVAARGRLMESARDDGAMLAIEA
PESAVTPLLDASVSLAAVNGPNAVVLAGDLERVLAIGAEFKAQGVRTKRLTVSHAFHSAHLDPMLDEYRR
IARGLSYSAPRIPVVSTVTGEVADELTSPEYWVRQVRQPVRFLDGMRTLYSSGVRTFVEIGPDAVLAPAMQA
CVPEVDHGVIPALTSRRSEPRTLVTALAKAHVRGVPVDWSAGFSGARRVDLPTYAFQRSRYWLDPEAPSAQ
VDLASAADVLGLTEVQRATLAAWQRRHTWAHRKTWRPLRTVPSPPGTWLVPSGPDRLVEVLRERGAQVV
TAGNAPVDGVLTFDEVDHPAPVWLLTSDADPGYACDHLVHLPTSPTDETWERLVDALGSAEPRLAIRGPE
VFAEHLERAELDGDWWPDGPVLITGDDELSSWAARWFEEQGVAVVAEAGPSLAAVVQVVPPEWTPAEVR
EALAPLDELTAMVFVFTSATRSPAQASATAWARRRGAVVV

QMN_KSATΔ21 (residues 1-1177 of M5)

MSTGPTEQKLLDYLKWVTTDLDRTRDQLLALEEKASEPIAVVAMSCRYPHGVTSPEELWELVASGRDAIA
DFPADRGWPADVYDPDPDRRDRPHAGGFLYDAADFDPGFFGISPREAPAIDPQQRLLLEIAWEAFERAGI
DAESVRGSDTGVFTGVMYGDYGSRLQDRAPQGYESYLGNGSAGSIASGRVAYTFGLEGPAVTIDTACSSSL
VALHLAVQALRQGDCSLALAGGVTVMATPMVFTEFDRQRGLSADGRCKSFAAAADGTGFAEGAGLLLL
ERLSDARRHGHPVLGLVRGTAVNQDGASNGLTAPNGPSQERVIKQALENARVTPSEVDAVEAHGTGTTL
GDPIEAQALLATYGKARGEKPLYLGSIKSNIGHTQAAAGVAGVIKMIMAMRHGELPRTLHVDAPNPLVD
WASGGVSLLTEPAKWQQNGHPRRAGVSSFGISGTNAHVILEQAPPVEPASEPCVAPGPVPWVLSAKSPAA
VQGQAERLLAHLTLNPGLTVADVGHSLATTRTSFSHRAVVFGDRQEELVKGLSGLAHGTPATNLVQGRAT
GGKTAFVFTGQGSQRPGMGRQLYATYPVFTAELDRVCALFDGLLDRPLLPILFAEPRTPEAALLNRTAYTQ
VALFALETALYRLVESYRVRPSFVAGHSIGELTAAHVAGVLSLEDACVLVAARGRLMESARDDGAMLAIEA
PESAVTPLLDASVSLAAVNGPNAVVLAGDLERVLAIGAEFKAQGVRTKRLTVSHAFHSAHLDPMLDEYRR
IARGLSYSAPRIPVVSTVTGEVADELTSPEYWVRQVRQPVRFLDGMRTLYSSGVRTFVEIGPDAVLAPAMQA
CVPEVDHGVIPALTSRRSEPRTLVTALAKAHVRGVPVDWSAGFSGARRVDLPTYAFQRSRYWLDPEAPSAQ
VDLASAADVLGLTEVQRATLAAWQRRHTWAHRKTWRPLRTVPSPPGTWLVPSGPDRLVEVLRERGAQVV
TAGNAPVDGVLTFDEVDHPAPVWLLTSDADPGYACDHLVHLPTSPTDETWERLVDALGSAEPRLAIRGPE
VFAEHLERAELDGDWWPDGPVLITGDDELSSWAARWFEEQGVAVVAEAGPSLAAVVQVVPPEWTPAEVR
EALAPLDELTAMVFVFTSATRSPAQASATAWARRRGAVVVVVVAWDGVAPSAVGLVLRQG

QMN M5_KSAT (residues 1-1198 of M5, Dr. Matthew Jenner)

Mutations were carried out on the KS active-site (C207A) and the AT active-site (S656C).

MSTGPTEQKLLDYLKWVTTDLDRTRDQLLALEEKASEPIAVVAMSCRYPHGVTSPEELWELVASGRDAIA
DFPADRGWPADVYDPDPDRRDRPHAGGFLYDAADFDPGFFGISPREAPAIDPQQRLLLEIAWEAFERAGI
DAESVRGSDTGVFTGVMYGDYGSRLQDRAPQGYESYLGNGSAGSIASGRVAYTFGLEGPAVTIDTA**C**SSSL
VALHLAVQALRQGDCSLALAGGVTVMATPMVFTEFDRQRGLSADGRCKSFAAAADGTGFAEGAGLLLL
ERLSDARRHGHPVLGLVRGTAVNQDGASNGLTAPNGPSQERVIKQALENARVTPSEVDAVEAHGTGTTL
GDPIEAQALLATYGKARGEKPLYLGSIKSNIGHTQAAAGVAGVIKMIMAMRHGELPRTLHVDAPNPLVD
WASGGVSLLTEPAKWQQNGHPRRAGVSSFGISGTNAHVILEQAPPVEPASEPCVAPGPVPWVLSAKSPAA
VQGQAERLLAHLTLNPGLTVADVGHSLATTRTSFSHRAVVFGDRQEELVKGLSGLAHGTPATNLVQGRAT
GGKTAFVFTGQGSQRPGMGRQLYATYPVFTAELDRVCALFDGLLDRPLLPILFAEPRTPEAALLNRTAYTQ
VALFALETALYRLVESYRVRPSFVAGH**S**IGELTAAHVAGVLSLEDACVLVAARGRLMESARDDGAMLAIEA
PESAVTPLLDASVSLAAVNGPNAVVLAGDLERVLAIGAEFKAQGVRTKRLTVSHAFHSAHLDPMLDEYRR
IARGLSYSAPRIPVVSTVTGEVADELTSPEYWVRQVRQPVRFLDGMRTLYSSGVRTFVEIGPDAVLAPAMQA
CVPEVDHGVIPALTSRRSEPRTLVTALAKAHVRGVPVDWSAGFSGARRVDLPTYAFQRSRYWLDPEAPSAQ
VDLASAADVLGLTEVQRATLAAWQRRHTWAHRKTWRPLRTVPSPPGTWLVPSGPDRLVEVLRERGAQVV
TAGNAPVDGVLTFDEVDHPAPVWLLTSDADPGYACDHLVHLPTSPTDETWERLVDALGSAEPRLAIRGPE
VFAEHLERAELDGDWWPDGPVLITGDDELSSWAARWFEEQGVAVVAEAGPSLAAVVQVVPPEWTPAEVR
EALAPLDELTAMVFVFTSATRSPAQASATAWARRRGAVVVVVVAWDGVAPSAVGLVLRQGAAAPDVVAEL
SASVPAPAVAP

QMN M5 (Dr. Orestis Lazos) KS-AT-ACP

Mutations were carried out on the KS active-site (C207A) and the AT active-site (S656C).

MSTGPTEQKLLDYLKWVTTDLDRTRDQLLALEEKASEPIAVVAMSCRYPHGVTSPEELWELVASGRDAIA
DFPADRGWPADVYDPDPDRRDRPHAGGFLYDAADFDPGFFGISPREAPAIDPQQRLLLEIAWEAFERAGI
DAESVRGSDTGVFTGVMYGDYGSRLQDRAPQGYESYLGNGSAGSIASGRVAYTFGLEGPAVTIDTA**C**SSSL
VALHLAVQALRQGDCSLALAGGVTVMATPMVFTEFDRQRGLSADGRCKSFAAAADGTGFAEGAGLLLL
ERLSDARRHGHPVLGLVRGTAVNQDGASNGLTAPNGPSQERVIKQALENARVTPSEVDAVEAHGTGTTL
GDPIEAQALLATYGKARGEKPLYLGSIKSNIGHTQAAAGVAGVIKMIMAMRHGELPRTLHVDAPNPLVD
WASGGVSLLTEPAKWQQNGHPRRAGVSSFGISGTNAHVILEQAPPVEPASEPCVAPGPVPWVLSAKSPAA
VQGQAERLLAHLTLNPGLTVADVGHSLATTRTSFSHRAVVFGDRQEELVKGLSGLAHGTPATNLVQGRAT

GGKTAFVFTGQGSQRPGMGRQLYATYPVFTAELDRVCALFDGLLDRPLLPILFAEPRTPEAALLNRTAYTQ
VALFALETALYRLVESYRVRPSFVAGHSIGELTAAHVAGVLSLEDACVLVAARGRLMESARDDGAMLAIEA
PESAVTPLLDASVSLAAVNGPNAVVLAGDLERVLAIGAEFKAQGVRTKRLTVSHAFHSAHLDPMLDEYRR
IARGLSYSAPRIPVVSTVTGEVADELTSPEYWVRQVRQPVRFLDGMRTLYSSGVRTFVEIGPDAVLAPAMQA
CVPEVDHGVIPALTSRRSEPRTLVTALAKAHVRGVPVDWSAGFSGARRVDLPTYAFQRSRYWLDPEAPSAQ
VDLASAADVLGLTEVQRATLAAWQRRHTWAHRKTWRPLRTVPSPPGTWLVPSGPDRLVEVLRERGAQVV
TAGNAPVDGVLTFDEVDHPAPVWLLTSDADPGYACDHLVHLPTSPTDETWERLVDALGSAEPRLAIRGPE
VFAEHLERAELDGDWWPDGPVLITGDDELSSWAARWFEEQGVAVVAEAGPSLAAVVQVVPPEWTPAEVR
EALAPLDELTAMVFVFTSATRSPAQASATAWARRRGAVVVVVVAWDGVAPSAVGLVLRQGAAAPDVVAEL
SASVPAPAVAPDNRLLAATGAEQEQLLTELVRFHASDVLGHASPEAIGVEDSLLDHGFSSFAALELSNRLQA
ASGVQIPPVAIYDQPTLGAVITHLRDALAAEPS

QMN M5_KR^{D}-A (residues 960-1198 of M5)

WLVPSGPDRLVEVLRERGAQVVTAGNAPVDGVLTFDEVDHPAPVWLLTSDADPG/YACDHLVHLPTSPT
DETWERLVDALGSAEPRLAIRGPEVFAEHLERAELDGDWWPDGPVLITGDDELSSWAARWFEEQGVAVV
AEAGPSLAAVVQVVPPEWTPAEVREALAPLDELTAMVFVFTSATRSPAQASATAWARRRGAVVVAWDGV
APSAVGLVLRQGAAAPDVVAELSASVPAPAVAP

QMN M5_KR^{D}-B (residues 1014-1198 of M5)

YACDHLVHLPTSPTDETWERLVDALGSAEPRLAIRGPEVFAEHLERAELDGDWWPDGPVLITGDDELSSW
AARWFEEQGVAVVAEAGPSLAAVVQVVPPEWTPAEVREALAPLDELTAMVFVFTSATRSPAQASATAWAR
RRGAVVVAWDGVAPSAVGLVLRQGAAAPDVVAELSASVPAPAVAP

QMN M5_KR^{D}-C (residues 1063-1198 of M5)

DGDWWPDGPVLITGDDELSSWAARWFEEQGVAVVAEAGPSLAAVVQVVPPEWTPAEVREALAPLDELT
AMVFVFTSATRSPAQASATAWARRRGAVVVAWDGVAPSAVGLVLRQGAAAPDVVAELSASVPAPAVAP

QMN M5_KR^{D}-D (residues 1103-1198 of M5)

LAAVVQVVPPEWTPAEVREALAPLDELTAMVFVFTSATRSPAQASATAWARRRGAVVVAWDGVAPSAVG
LVLRQGAAAPDVVAELSASVPAPAVAP

QMN M3_^{ACP}dd-**M5_dd**^{KS}

APEPTDELDRLAADLSTMDPERQAAVLGRMAELLHKWNTNGTGNGASADLDSLSDEELFD
ALDEELGQRGNGRR**MSTGPTEQKLLDYLKWVTTDLDRTRDQLLALEEKASEP**

QMN M4_^{ACP}dd-**M5_dd**^{KS}

APRTEEPDPASAVLERLAELEEALAGLSDPRIADRLRELLAVAGGTDPGLGSASADELFD
FIDGKLGRKVETA**MSTGPTEQKLLDYLKWVTTDLDRTRDQLLALEEKASEP**

Known genes from the QMN biosynthetic cluster

Table 8.1. *The QMN gene cluster and the proposed functions of encoded proteins.*

Gene name	Length (bp)	Proposed function of encoded protein
***qmn*B**	1,413	propionyl-CoA carboxylase
***qmn*C**	753	TE II
***qmn*D$_4$**	966	2-oxoacid dehydrogenase/AT
***qmn*A$_1$**	17,772	PKS (KS^Q-AT-ACP-KS-AT-DH-KR-ACP-KS-AT-DH-KR-ACP-KS-AT-DH-KR-ACP)
***qmn*A$_2$**	5,313	PKS (KS-AT-DH-KR-ACP)
***qmn*A$_3$**	3,876	PKS (KS-AT-KR^D-ACP)
***qmn*E**	1,236	unknown
***qmn*F**	1,488	unknown
***qmn*Rg$_1$**	255	regulator
***qmn*G**	1,599	PQQ-dependent dehydrogenase
***qmn*H**	1,128	Proposed Diels Alderase
***qmn*I**	1,044	unknown
***qmn*J**	483	unknown
***qmn*Rs$_1$**	1,176	transporter
***qmn*Rs$_2$**	681	transporter
***qmn*K**	1,056	unknown
***qmn*L**	456	unknown
***qmn*M**	657	unknown
***qmn*N**	390	unknown
***qmn*Rg$_2$**	1,347	regulator
***qmn*Rg$_1$**	666	regulator
***qmn*D$_5$**	1,029	3-oxoacyl-ACP synthase III (KS)
***qmn*D$_1$**	1,839	glyceryltransferase/phosphatase
***qmn*D$_2$**	213	ACP
***qmn*D$_3$**	843	2-oxoacid dehydrogenase/AT
***qmn*O**	1,209	cytochrome P450
***qmn*Rg$_4$**	2,805	regulator
***qmn*Rg$_5$**	2,463	regulator

NMR and mass spectra for the production of decanoyl- and dodeconoyl-pantetheine

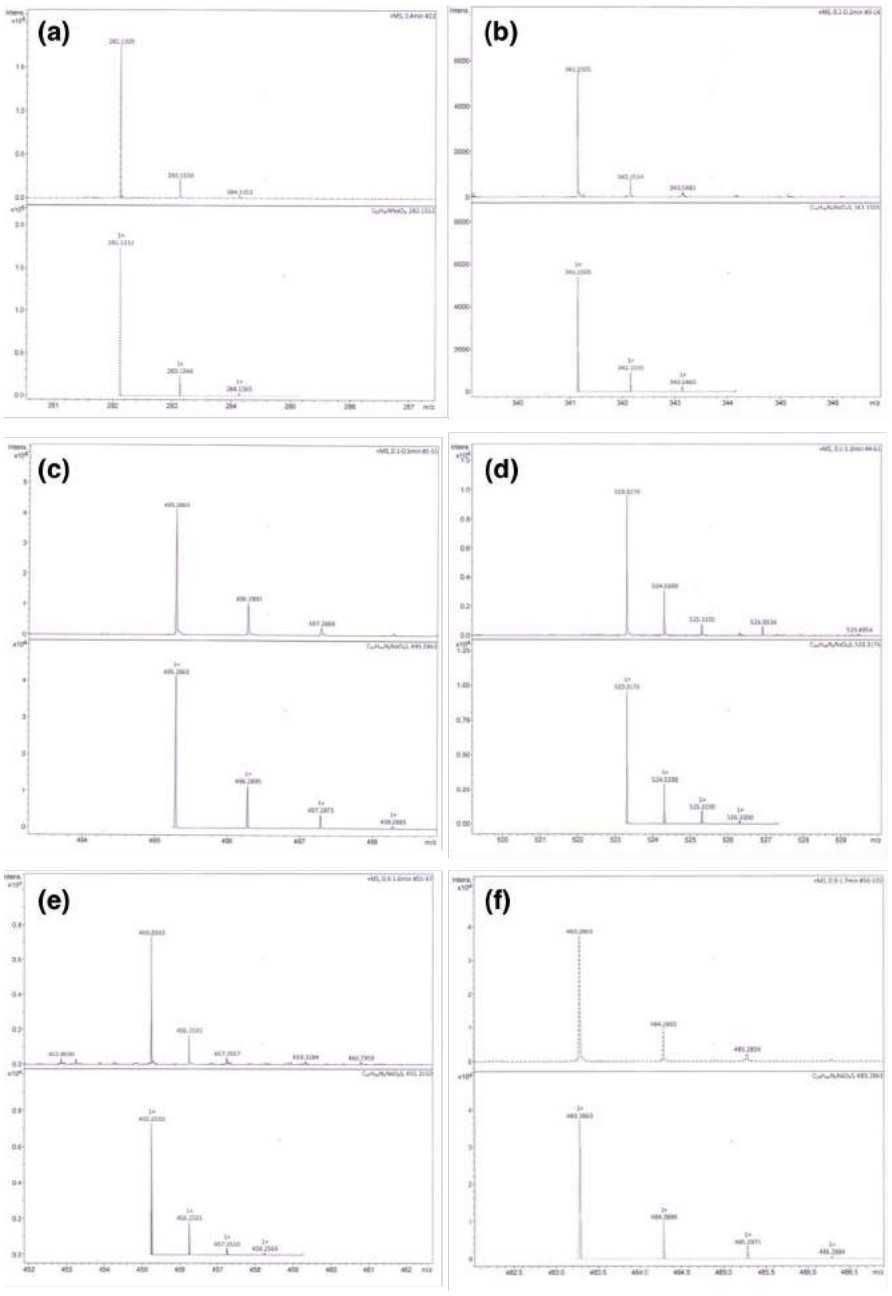


Figure 8.1. *Mass spectra of each step in decanoyl- and dodecanoyl-pantetheine synthesis.* ***(a)*** *Protected pantothenic acid,* ***(b)*** *protected pantetheine,* ***(c)*** *protected decanoyl pantetheine,* ***(d)*** *protected dodecanoyl-pantetheine,* ***(e)*** *deprotected decanoyl-pantetheine,* ***(f)*** *deprotected dodecanoyl-pantetheine.*

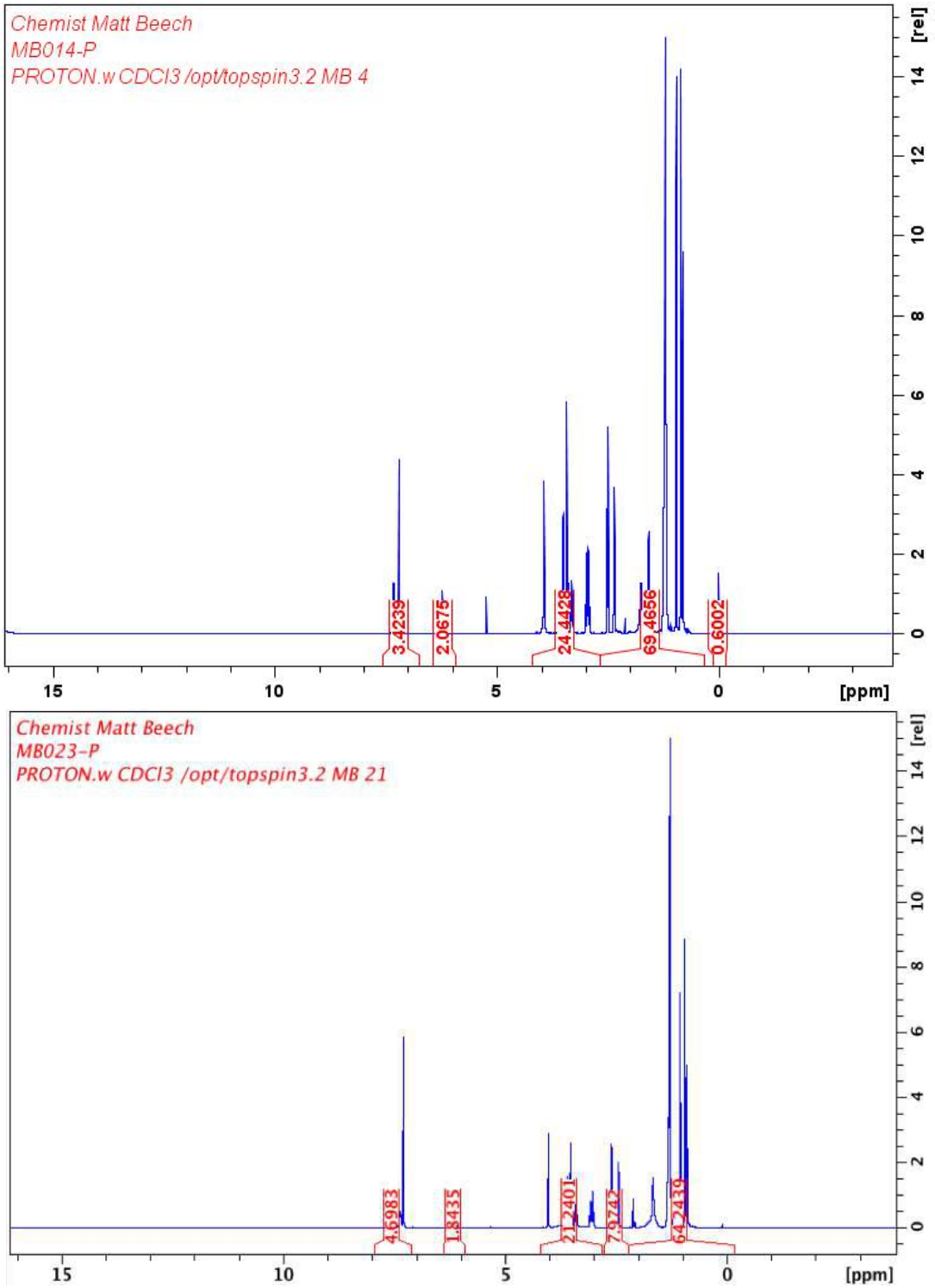
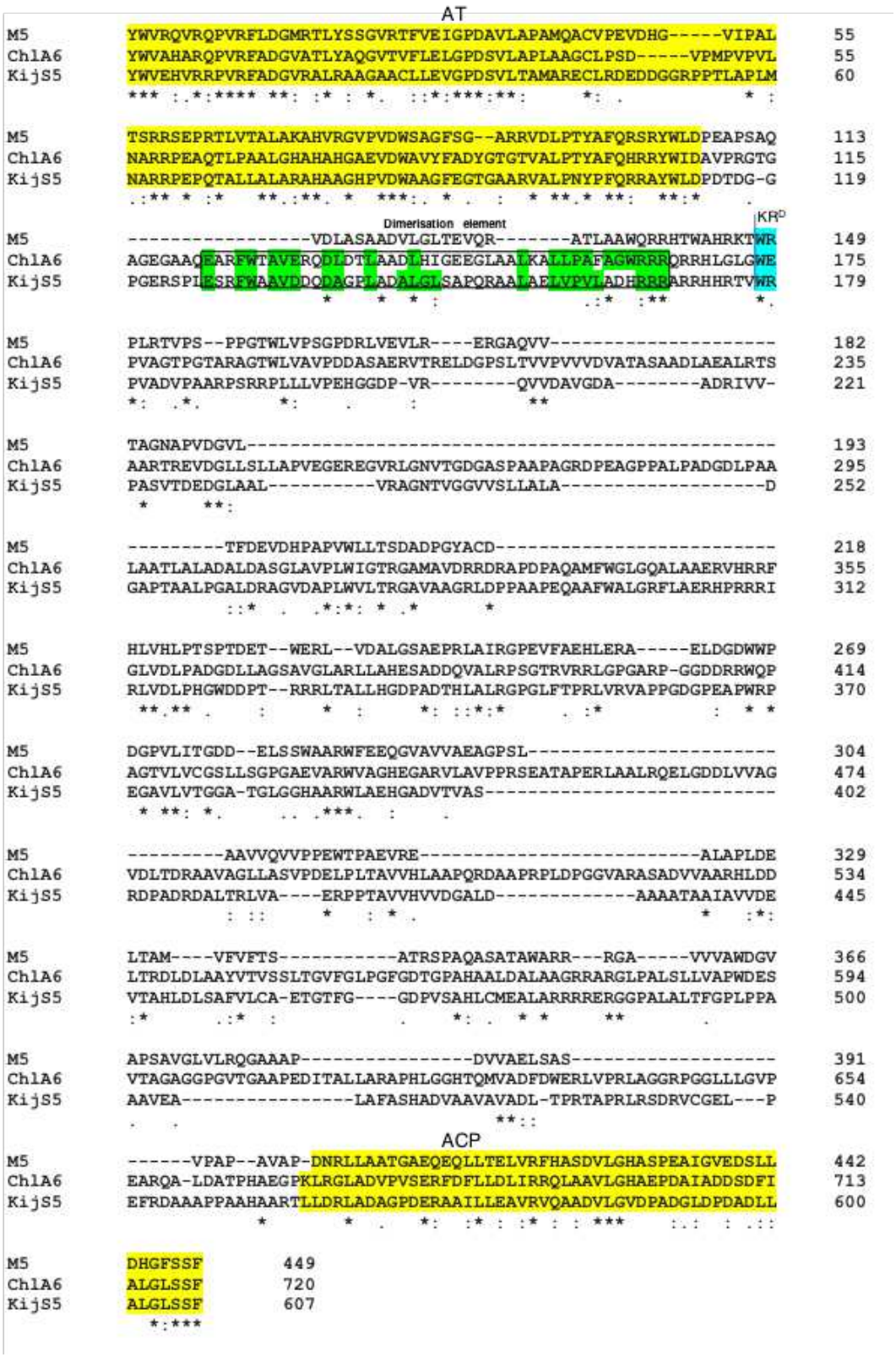


Figure 8.2. ^{1}H *NMR analysis of decanoyl- (above) and dodecanoyl-pantetheine (below).*

Sequence alignments



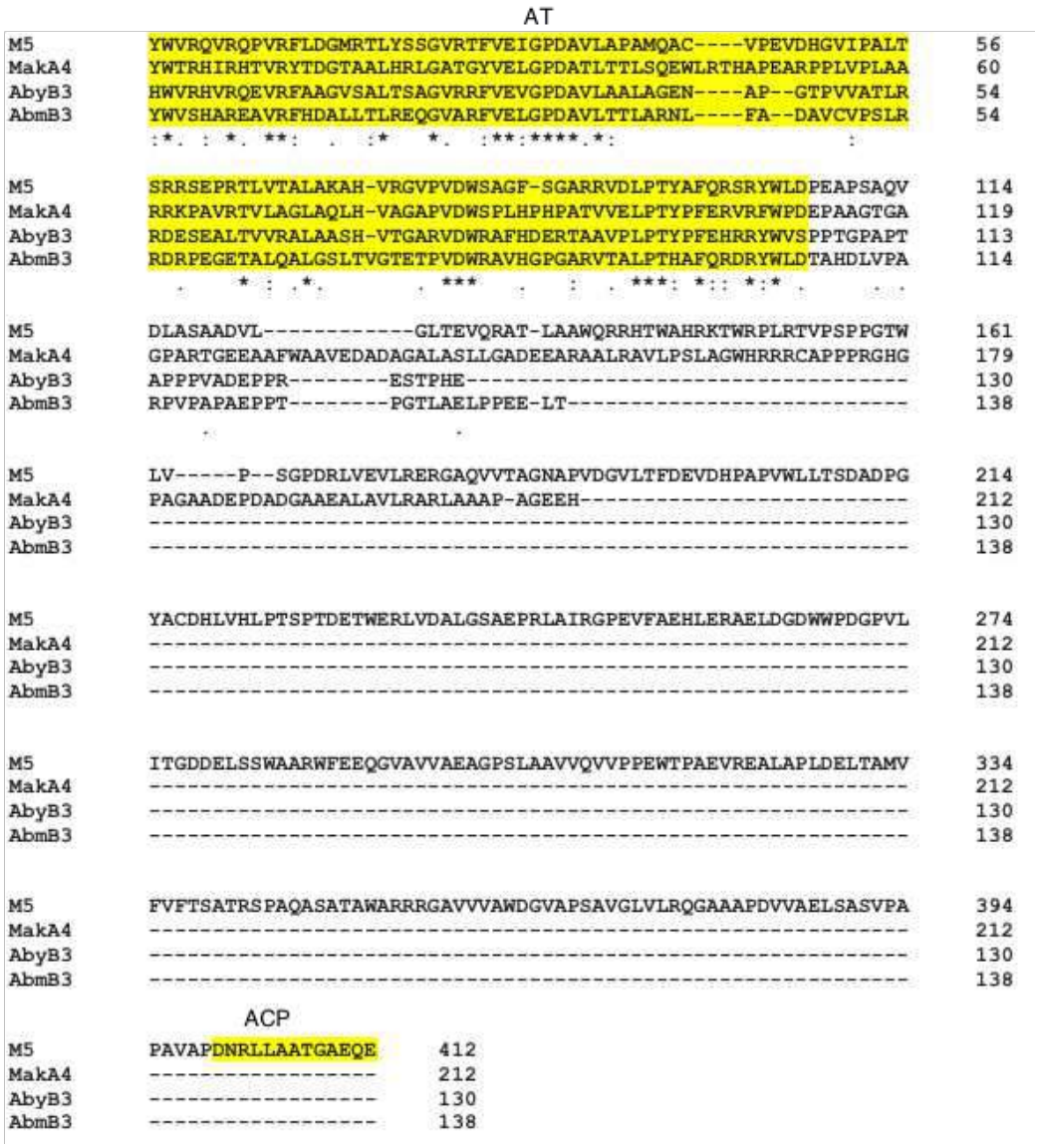


Figure 8.3. *Sequence alignment of QMN M5's linker region with other AT-ACP regions of spirotetronate PKS final modules.*

Mass spectra

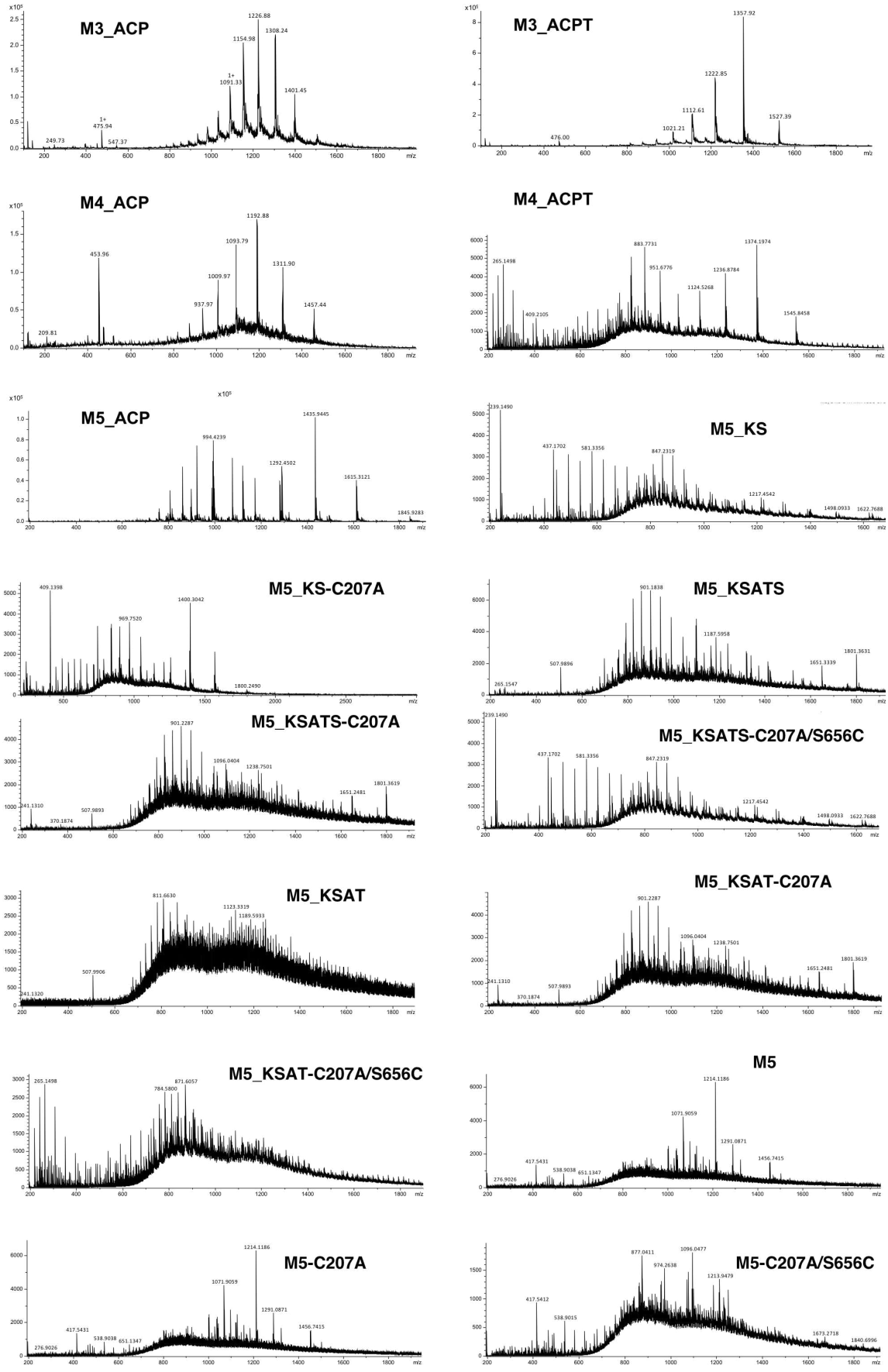


Figure 8.4. *Mass spectra for overproduced proteins characterised in chapter 2.*